AD-755 918

RESEARCH ON DIAGNOSTIC EVALUATION OF SPEECH INTELLIGIBILITY

William D. Voiers, et al

TRACOR, Incorporated

Prepared for:

Air Force Cambridge Research Laboratories

24 January 1973

DISTRIBUTED BY:

NTIS

National Technical Information Service
U. S. DEPARTMENT OF COMMERCE
5285 Port Royal Road, Springfield Va. 22151

AD755918

AFCRL-72-0694

RESEARCH ON DIAGNOSTIC EVALUATION OF SPEECH INTELLIGIBILITY

by

William D. Voiers,

Alan D. Sharpley,

and

Carl J. Hehmsoth

TRACOR, Inc., 6500 Tracor Lane, Austin, Texas 78721

Contract No. F 19628-70-C-0182

Project No. 4610
Task No. 461002
Work Unit No. 46100201

FINAL REPORT

1 February 1970 - 31 August 1972

24 January 1973

Contract Monitor: Caldwell P. Smith
Data Sciences Laboratory

Approved for public release; distribution unlimited

Prepared
for

AIR FORCE CAMBRIDGE RESEARCH LABORATORIES
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
BEDFORD, MASSACHUSETTS 01730

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Tracor, Inc. 6500 Tracor Lane Austin, Texas 78721	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

RESEARCH ON DIAGNOSTIC EVALUATION OF SPEECH INTELLIGIBILITY

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Scientific Final. 1 February 1970 - 31 August 1972 Approved: 19 Dec., 1972

5. AUTHOR(S) (First name, middle initial, last name)

William D. Voiers
Alan D. Sharpley
Carl J. Hehmsoth

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
30 September 1972	138	32

8a. CONTRACT OR GRANT NO. F19628-70-C-0182 b. PROJECT NO. 4610-02-01 c. DOD ELEMENT 62702F d. DOD SUBELEMENT 674610	9a. ORIGINATOR'S REPORT NUMBER(S) T-73-AU-9010-U 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) AFCRL-72-0694

10. DISTRIBUTION STATEMENT

A- Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES TECH, OTHER	12. SPONSORING MILITARY ACTIVITY Air Force Cambridge Research Laboratories (LR) L. G. Hanscom Field, Bedford, Mass. 01730

13. ABSTRACT

The activities of the Psychometrics Department of TRACOR, Inc., under Contract No. F 19628-70-C-0182, fall into two major categories. In the first category are research activities undertaken with the aim of developing improved methods for evaluating voice communication systems and devices. In the second category are testing services performed with processed speech materials supplied by the contract monitor. The research activities included five major research projects from which technical papers resulted. They are: "Diagnostic Approach to the Evaluation of Speech Intelligibility," "The Nature of Individual Differences in Diagnostic Rhyme Test Performance," "Speaker Effects on Intelligibility Test Results," "Structure of Phonemic Information in the Oral and Nasal Outputs," "Diagnostic Evaluation of Intelligibility of Present-day Digital Vocoders" (jointly authored by W. D. Voiers and C. P. Smith). These papers constitute Chapters 1-5 of this report. A summary of activities constitutes Chapter 6. The results of diagnostic intelligibility tests were forwarded to the contract monitor as they became available during the period of the contract.

DD FORM 1 NOV 65 1473

Unclassified
Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Speech Intelligibility						
Diagnostic Rhyme Test						
Speaker Factors						
Listener Factors						

ib

AFCRL-72-0694

RESEARCH ON DIAGNOSTIC EVALUATION OF SPEECH INTELLIGIBILITY

by

William D. Voiers,

Alan D. Sharpley,

and

Carl J. Hehmsoth

TRACOR, Inc., 6500 Tracor Lane, Austin, Texas 78721

Contract No. F 19628-70-C-0182

Project No. 4610
Task No. 461002
Work Unit No. 46100201

FINAL REPORT

1 February 1970 - 31 August 1972

24 January 1973

Contract Monitor: Caldwell P. Smith
Data Sciences Laboratory

Approved for public release; distribution unlimited

Prepared
for

AIR FORCE CAMBRIDGE RESEARCH LABORATORIES
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
BEDFORD, MASSACHUSETTS 01730

ic

ABSTRACT

The activities of the Psychometrics Department of TRACOR, Inc., under Contract No. F 19628-70-C-0182 fall into two major categories. In the first category are research activities undertaken with the aim of developing improved methods for evaluating voice communication systems and devices. In the second category are testing services performed with processed speech materials supplied by the contract monitor.

The research activities included five major research projects from which technical papers resulted. They are:

"Diagnostic Approach to the Evaluation of Speech Intelligibility"

"The Nature of Individual Differences in Diagnostic Rhyme Test Performance"

"Speaker Effects on Intelligibility Test Results"

"Structure of Phonemic Information in the Oral and Nasal Outputs"

"Diagnostic Evaluation of Intelligibility of Present-Day Digital Vocoders" (jointly authored by W. D. Voiers and C. P. Smith)

These papers constitute Chapters 1-5 of this report. A summary of activities constitutes Chapter 6. The results of diagnostic intelligibility tests were forwarded to the contract monitor as they became available during the period of the contract.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF TABLES (cont.)

CHAPTER 1

DIAGNOSTIC APPROACH TO THE EVALUATION OF SPEECH INTELLIGIBILITY

by

William D. Voiers

PREFACE

During the three years in which the Diagnostic Rhyme Test Form III was used for purposes of research and system evaluation, a mass of data bearing upon the intrinsic difficulty of individual test items was accumulated. Examination of these data revealed various indications that variation in difficulty among items of a given type is, at least in part, of systematic origin.

There were several indications that item difficulty varies with vowel context. More pronounced, however, were indications that the apprehensibility of a given feature varies with the states of other features in the same phoneme. This phenomenon is termed an ipsative dependency to distinguish it from the types of transitive dependencies usually referred to as coarticulation effects. For example, of the items designed to test for the apprehensibility of voicing, those items in which the critical phonemes are sustained (e.g., /v/ and /f/) appeared to be more difficult generally than items in which the critical phonemes are interrupted (e.g., /b/ and /p/). Among the items designed to test the apprehensibility of graveness, items which involved unvoiced critical phoneme pairs appeared under some conditions (e.g., noise-masked and low passed speech) to be more difficult than items involving voiced pairs. The reverse of this trend was observed, however, in the case of high passed speech, and there were other instances of dependencies which appeared to be interactive with the transmission condition involved.

Such dependencies are clearly of potential diagnostic significance, but while they can be detected with Form III of the DRT, their evaluation is, in most instances, a rather cumbersome process. It was clearly desirable, therefore, to design a test, the structure of which would permit relatively rigorous statistical evaluation of both ipsative and transitive dependencies. Accordingly, modification of the DRT was undertaken to the end of providing a test in which various dependencies of both types would be amenable to routine statistical evaluation. The culmination of this effort was Diagnostic Rhyme Test Form IV (DRT IV), which is described in the following report.

INTRODUCTION

It is a matter of common observation that speech communication -- more specifically, a listener's apprehension of a speaker's linguistic intent -- is essentially a dual process. One aspect of this process, the perceptual aspect, involves discriminations by the listener of various acoustical manifestations of the speaker's intent. The other, apperceptual, aspect involves inferences based on contextual or extra-stimulus information, i.e., on information from sources extrinsic to the immediate acoustical correlates of the speaker's intent. Thus the listener's uncertainty with regard to a speaker's intent may be reduced by such factors as his knowledge of the structure of the language involved;[1,2] his knowledge of the circumstances occasioning and the purposes motivating the communication;[3] his familiarity with dialectal and idiolectal characteristics of the speaker;[4] and his knowledge of the immediate past history of the speech signal.[5]

Both the perceptual and the apperceptual aspects of the speech apprehension process are legitimate subjects of scientific interest. For most scientific purposes, however, it is essential that they be subject to independent experimental control. Clearly, it is essential that contextual effects be controlled in listening tests conducted to evaluate the intrinsic characteristics of a transmission channel or medium as well as in experiments concerned with certain aspects of the processes of speech production and perception. To the extent that a listener's responses in the testing situation are dependent to an unknown degree upon contextual information, his performance necessarily provides an imperfect reflection of the entity or process under evaluation. Although cognizance of this issue is at least implicit in the designs of most speech reception tests in use today, a number of problems remain. These problems become particularly acute, moreover, in those instances where some form of "diagnostic" scoring is to be attempted, i.e., where significance is to be attributed not only to the number, but also to the types of errors committed by the listener.

Among the more formidable problems complicating the design and use of speech reception tests is the problem of controlling the effects of the listener's familiarity with the test materials used, and a variety of procedures have been devised to cope with it. In the case of the Harvard Phonetically Balanced (PB) Test, for example, the recommended procedure for controlling familiarity involves an extensive regimen of training, terminated on evidence that the effects of familiarity have reached an asymptotic state. This approach to the problem serves, among other things, to limit the circumstances in which use of the "PB" test is practical. More crucial, however, are its potential effects on the validity of results obtained with the test.

Familiarization training serves most immediately to alter the general level of difficulty of the listener's task, and thus to obscure any relationship between the "real world" and the testing situation that might be claimed on the basis of absolute level of difficulty. Additionally, however, familiarization training may effect qualitative changes in the listener's task and thus in the implications of his performance. This possibility derives from the fact that the various discriminations required of the listener in the course of recognizing a speech sound are not of intrinsically equal difficulty. As shown by Miller and Nicely, for example, some discriminations are accomplished with virtually perfect reliability, even under conditions of extreme signal impoverishment. Others are accomplished with significantly less than perfect reliability under the best of conditions, and may become prohibitively difficult under conditions of signal impoverishment. In view of these considerations, it would seem to be an extremely tenuous assumption that the facilitative effects of familiarization training are exerted equally on all aspects of the speech discrimination task. The alternative possibility is that familiarization training facilitates listener performance primarily in the more difficult aspects of the speech discrimination task. Effectively, therefore, it may desensitize the test primarily with respect to the acoustic speech features most crucial to the communication process and, perhaps, most vulnerable to common forms of signal impoverishment. In any case, the inter-phonemic constraints characteristic of the "PB" and similar word recognition tests preclude any type of qualitative or "diagnostic" evaluation of listener errors. Such constraints hopelessly confound the effects of contextual factors with effects attributable to the characteristics of the entity under test.

Testing procedures in which stimulus uncertainty is limited to a single phoneme (as in the Fairbanks Rhyme Test[8]), and particularly where the listener's response options are explicitly specified (as in the Modified Rhyme Test [9] and the Phonemically Balanced Rhyme Test[10]), substantially reduce the effects of familiarity upon listener performance. However, restriction of the listener's reponse options, whether implicit or explicit, may complicate the interpretation of test results in other ways, particularly if significance is to be attributed to the type as well as to the number of errors committed by the listener. To restrict the listener's response options in an arbitrary or unsystematic manner may be to substitute one set of unknown contextual constraints for another, such that stimulus effects upon the type of error committed become confounded with contextual effects. The crucial point here is that the discriminations required of a listener in identifying a complex stimulus are determined not by the characteristics of the stimulus as regarded in isolation, but rather by the characteristics that distinguish the stimulus from what the listener conceives to be the set of possible stimuli in a given instance. Thus, to constrain the listener's options in an unsystematic manner is possibly to deny him opportunities for providing information concerning the discriminability of certain speech features. This, in turn, may serve to desensitize a speech reception test with respect to specific deficiencies of the communication system, speaker, or listener being tested.

The hazards of restricted response sets can be minimized by means of carefully designed test items, particularly where the differences between the correct and incorrect response options are in some sense univocal. For example, in the ensemble:

bee pea vee dee me,

each permissible, erroneous response differs from the correct response, "bee," by a single "distinctive feature." Tests composed of such items could be quite effective in circumstances where the individual listener does not experience repeated exposure to the test materials.

Problems arise, however, where it is desirable to have different, but equivalent randomizations of multiple choice test materials. If, for example, "pea" were the stimulus word in the above ensemble, "unidimensional" differences between correct and incorrect options would no longer obtain. Only "bee" differs minimally and unidimensionally from the stimulus word, while other options differ by two or more "distinctive features." The structure of a test composed of such items would thus tend to vary somewhat with different randomizations of the test materials and to greatly complicate the mechanics of both gross and diagnostic scoring.

From the foregoing it is evident that the multiple choice approach, in general, has certain limitations as well as intrinsic advantages. Many of these limitations can be overcome by recourse to the special case of two-choice testing procedures. With such procedures, erroneous responses can, but for the effects of chance, be attributed unequivocally to the characteristics of the entity under test. Because of the inherent redundancy of the speech signal, however, phonemic confusion data do not ordinarily suffice for exact specification of deficiencies of the system or other entity under test. Rather, a phonemic confusion implies a deficiency in the encoding, transmission or discrimination of _one or more_ acoustical speech features, the precise number and nature of which cannot be specified without additional information.

Given this circumstance, it is clearly desirable at least to minimize uncertainty as to the feature or features involved. The means to this end is provided by a phonemic taxonomy broadly patterned after the distinctive feature systems of Jacobson, Fant and Halle,[11] Miller and Nicely,[12] and De Lattre.[13] Such a taxonomy provides a basis for the construction of a two-choice test where the correctness of the listener's response to a given item is criterial -- depending on the design and purposes of the investigation -- of the effective fidelity with which a speaker articulates, a system transmits, or the listener himself can discriminate the states of a limited set, or cluster, of intercorrelated, information-bearing, acoustical features. Data yielded by such a test can serve to sharply delimit the possible sources of deficiency or malfunction in an entity under test, and may serve in conjunction with other information to identify, precisely, source of malfunction or deficiency.

The Diagnostic Rhyme Test (DRT), in all of its versions, was designed on the basis of the foregoing considerations. Accordingly, it is a two-choice test in which each item involves two rhyming words, the initial consonants of which differ by a single phonemic attribute or feature. The listener's task is simply to judge which of the two words has been spoken, indicating, in effect, that he has or has not apprehended the speaker's intent with respect to the state of a particular phonemic attribute.

In addition to the theoretical advantages that can be realized with a two-choice approach, there are some significant practical advantages. Among them are: (1.) economy of testing time and materials, in that the use of minimally contrasting word pairs serves to exclude excessively easy and, hence, effectively nonfunctional items; (2.) minimal requirements with regard to listener selection and training (previous experience with the test materials can serve to facilitate listener performance only with respect to a particular randomization of the test materials); (3.) adaptability to both manual and computer scoring schemes; (4.) ease with which structurally equivalent randomizations can be generated.

Table 1 presents the phonemic taxonomy used as a basis for the design of the DRT, in which the six dimensions: voicing, nasality, sustention, sibilation, graveness, and compactness are represented. No provision is made to test apprehensibility of "vowel likeness," but constraints are observed in item construction to prevent covariation of this attribute with any of the above.

The articulatory and acoustical correlates of the phonemic attributes (or their equivalents in other classification systems) with which the DRT is concerned are extensively described in the recent literature.[14] Only the more important of these are indicated in Table 2. In accordance with the principle that consonant phonemes carry the bulk of the useful information in speech, and are also most susceptible to degradation, the scope of the DRT, like the Fairbanks Rhyme Test and its derivatives, is concerned only with the apprehensibility of consonants. Also like the FRT, the DRT treats only the case of consonant apprehension in the initial position. Although it is recognized that consonants may

TABLE 1. Consonant Taxonomy Used in the Construction of the DRT (Form IV).

	/m/	/n/	/v/	/ð/	/z/	/ʒ/	/ʒ̂/	/b/	/d/	/g/	/w/	/r/	/l/	/j/	/f/	/θ/	/s/	/ʃ/	/ʃ̂/	/p/	/t/	/k/	/h/
Voicing	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-
Nasality	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sustention	-	-	+	+	+	+	-	-	-	-	+	+	+	+	+	+	+	+	-	-	-	-	+
Sibilation	-	-	-	-	+	+	+	-	-	-	-	-	-	-	-	-	+	+	+	-	-	-	-
Graveness	+	-	+	-	-	o	o	+	-	o	+	-	o	o	+	-	-	o	o	+	-	o	o
Compactness	-	-	-	-	-	+	+	-	-	+	-	-	o	+	-	-	-	+	+	-	-	+	+
Vowel-like*	-	-	-	-	-	-	-	-	-	-	+	+	+	+	-	-	-	-	-	-	-	-	-

* The DRT does not test for the apprehensibility of the opposition, vowel-like - nonvowel-like. However, test words are chosen so as not to confound this attribute with the six attributes for which discriminability is tested.

TABLE 2. Major Genetic and Acoustical Correlates of Six Consonant Attributes.

		Genetic Correlates	Acoustical Correlates
Voicing	Present	Breath stream modulated by quasi-periodic vibration of vocal cords.	Harmonic structure evident; early onset of low frequency component.
	Absent	Vocal cords stationary.	Harmonic structure lacking.
Nasality	Present	Velum lowered.	Energy concentration in region of 250 Hz, "fast-reverse" formant transitions.
	Absent	Velum raised.	"Discontinuity of links."
Sustention	Present	Vocal tract not closed.	Gradual onset of energy; duration usually greater than 130 msec.
	Absent	Vocal tract completely closed.	Abrupt onset of energy; duration usually less than 130 msec.
Sibilation	Present	Rough-edged obstruction.	High intensity, high frequency noise.
	Absent	Smooth-edged obstruction.	Lower intensity, negligible noise.
Graveness (Initial position)	Present	Anterior articulation.	Relatively low loci of second and third formants.
	Absent	Medial articulation.	Relatively high loci of second and third formants.
Compactness (Initial position)	Present	Posterior articulation.	Divergence of second and third formants.
	Absent	Anterior or medial articulation.	Second and third formants convergent or parallel.

be differentially perceptible in the initial, intervocalic and terminal positions, the features involved are assumed to be equally represented in all positions.

THE DIAGNOSTIC RHYME TEST (DRT)

Structure of the DRT

The Diagnostic Rhyme Test (DRT) is more properly described in terms of a set of principles for item construction and selection than in terms of a specific corpus of test materials. Thus, the corpus of 96 rhyming word pairs shown in Table 3 constitutes only one realization of such principles, but takes into account the results of various experimental investigations conducted with earlier versions of the DRT. The gross structure of the test is evident in the table, where the items in each block of seven are arranged according to the attribute involved. The order is as follows:

1. Voicing
2. Nasality
3. Sustention
4. Sibilation
5. Graveness
6. Compactness
7. Filler item (to be used for research purposes, etc.)

The positive state (e.g., grave) of each attribute is represented in the left member of each pair; the negative state (e.g., acute) is represented in the right member of each pair.

The apprehensibility of each attribute is tested in each of

TABLE 3. Speech Materials Used in Form IV of the Diagnostic Rhyme Test.

99.*	VEAL-FEEL	43.	BEAN-PEEN	50.	ZOO-SUE	106.	DUNE-TUNE
107.	MEAT-BEAT	51.	NEED-DEED	2.	MOOT-BOOT	58.	NEWS-DUES
59.	VEE-BEE	3.	SHEET-CHEAT	66.	FOO-POOH	10.	SHOES-CHOOSE
67.	ZEE-THEE	11.	CHEEP-KEEP	74.	JUICE-GOOSE	18.	CHEW-COO
19.	WEED-REED	75.	PEAK-TEAK	82.	MOON-NOON	26.	POOL-TOOL
27.	YIELD-WIELD	83.	KEY-TEA	34.	COOP-POOP	90.	YOU-RUE
35.**	-----	91.**	-----	98.**	-----	42.**	-----
71.	GIN-CHIN	15.	DINT-TINT	22.	VOLE-FOAL	78.	GOAT-COAT
79.	MITT-BIT	23.	NIP-DIP	30.	MOAN-BONE	86.	NOTE-DOTE
31.	VILL-BILL	87.	THICK-TICK	38.	THOSE-DOZE	94.	THOUGH-DOUGH
95.	JILT-GILT	39.	SING-THING	46.	JOE-GO	102.	SOLE-THOLE
47.	BID-DID	103.	FIN-THIN	110.	BOWL-DOLE	54.	FORE-THOR
55.	HIT-FIT	111.	GILL-DILL	6.	GHOST-BOAST	62.	SHOW-SO
7.**	-----	63.**	-----	70.**	-----	14.**	-----
8.	ZED-SAID	64.	DENSE-TENSE	57.	VAULT-FAULT	1.	DAUNT-TAUNT
72.	MEND-BEND	16.	NECK-DECK	65.	MOSS-BOSS	9.	GNAW-DAW
80.	THEN-DEN	24.	FENCE-PENCE	17.	THONG-TONG	73.	SHAW-CHAW
32.	JEST-GUEST	88.	CHAIR-CARE	81.	JAWS-GAUZE	25.	SAW-THAW
40.	MET-NET	96.	PENT-TENT	33.	FOUGHT-THOUGHT	89.	BONG-DONG
104.	KEG-PEG	48.	YEN-WREN	97.	YAWL-WALL	41.	CAUGHT-TAUGHT
56.**	-----	112.**	-----	49.**	-----	105.**	-----
36.	VAST-FAST	92.	GAFF-CALF	85.	JOCK-CHOCK	29.	BOND-POND
44.	MAD-BAD	100.	NAB-DAB	93.	MOM-BOMB	37.	KNOCK-DOCK
52.	THAN-DAN	108.	SHAD-CHAD	101.	VON-BON	45.	VOX-BOX
4.	JAB-GAB	60.	SANK-THANK	109.	JOT-GOT	53.	CHOP-COP
12.	BANK-DANK	68.	FAD-THAD	61.	WAD-ROD	5.	POT-TOT
76.	GAT-BAT	20.	SHAG-SAG	69.	HOP-FOP	13.	GOT-DOT
84.**	-----	28.**	-----	21.**	-----	77.**	-----

* Numbers to the left of each pair indicate the position of the item in each block of 112 items on the listeners' answer sheet.

** Filler items. The manner in which these spaces are filled is at the option of the experimenter. Among other things, they may be used for testing experimental items.

eight vowel contexts. This involves two vowels from each "quadrant" of the vowel articulation diagram. Thus the four upper left blocks of Table 3 involve high, front vowels, whereas those in the four upper right blocks involve high, back vowels. The low, front vowels are represented in the four lower left blocks, while the low, back vowels are represented in the lower right blocks. No central vowels are used in the DRT.

There are two grossly equivalent items (e.g., bean-peen and veal-feel) designed to test for the apprehensibility of each attribute in each vowel context, which redundancy serves, among other things, to facilitate various tests of the reliability or consistency of listener performance over the course of a testing session. Either member of each pair may be chosen as the stimulus word in a given instance without changing the function of the item qualitatively. Choice of stimulus word affects only the polarity of the test provided by the item.

It is perhaps apparent from the table that insufficient latitude exists to permit any degree of selectivity on the basis of frequency of word occurrence in speech or printed matter. However, results such as those of Pollack, Rubinstein and Decker[15] suggest that frequency of use influences the perceptibility of complex stimuli primarily, if not only, as it provides a basis for the listener's expectation concerning the occurrence of the stimulus. Where other, more explicit, bases for expectation are available -- as they are in the case of the DRT -- frequency of use may reasonably be expected to have little or no influence on listener response, particularly, perhaps, where the listener is required, in effect, simply to discriminate a specific aspect of the total stimulus event, rather than to "recognize" the stimulus.

It may also be noted by reference to Tables 2 and 3 that there are some minor exceptions to the rule of "unidimensional difference" between members of each word pair. This results from the fact that all compact items are here classified indifferently with respect to graveness (rather than positively, as in Halle's taxonomy).[16] Thus, while the phonemes comprising the pairs /k-p/, /g-b/, /k-t/, /g-d/, etc., differ primarily with respect to compactness, they might be considered to differ secondarily in terms of graveness in that the first member of each pair has a neutral or indeterminate status with respect to the latter attribute, while the second member of each pair has a positive or negative status. In terms of the taxonomy in Table 2, there are, in other words, no phoneme pairs whose members are distinguished purely on the basis of compactness. However, adoption of Halle's system, whatever its merits in this application, would restrict the available phoneme pairs to those involving the "back-front" opposition. Data on phonemic confusability (e.g., Miller and Nicely)[17] suggest that the solution proposed here tends to conform most nearly with the facts of phonemic perception. Some experimental justification for this course of action is also provided by results to the effect that the apprehensibility of compactness, as measured by such items, is quite differently affected by various forms of signal impoverishment than is graveness.

In recognition of experimental evidence that the acoustical correlates of the state of a given attribute may not be equally apprehensible in every instance of its manifestation, nor equally vulnerable to all forms of signal impoverishment, various additional constraints were imposed in assembling the corpus shown in Table 3. Among the more important of these are:

1. In one-half the items designed to test for the apprehensibility of voicing, both critical phonemes involve friction; in the other half, friction is absent. Balance between the upper and lower and between the front and back halves of the vowel space is maintained with respect to these three taxonomic dimensions as well as to graveness and compactness.
2. Half of the nasality items in each vowel context lie in the "grave plane," i.e., involve grave phoneme pairs; half are in the acute plane. All, of course, lie at the intersection of the voiced, interrupted, unsibilated, and diffuse planes.
3. Half of the items designed to test for the apprehensibility of sustention lie in the voiced plane; half in the unvoiced. This dichotomy is not preserved within each vowel

context, due to the constraints inherent in the language, but each quadrant of the vowel space is balanced in this respect.

4. Half of the <u>sibilation</u> items in each vowel context lie in the voiced plane; half lie in the unvoiced. But for the pair ZEE-THEE, there is perfect symmetry of halves of the vowel space.

5. In the case of <u>graveness</u>, items were selected such that, for each vowel environment, one item lies in the voiced plane, one in the unvoiced; one lies in the sustained plane, one in the interrupted.

6. In addition to the constraints previously noted with respect to <u>compactness</u>, items were selected such that, for each quadrant of the vowel space, one item lies in the vowel-like plane and one item lies in the sibilated plane. All combinations of the states of <u>voicing</u> and <u>sustention</u> are given equal representation in each quadrant of the vowel plane.

With minor exceptions, the two halves of the vowel space, partitioned horizontally or vertically, involve identical phoneme pairs for testing the apprehensibility of any attribute.

<u>Preparation of Stimulus Materials</u>

The first steps in the preparation of test speech materials involve the determination of sequential arrangements of items and the selection of a stimulus word from each item. Assuming that adequate precautions are made to counterbalance the effects of fatigue, "warm up," etc., there are no theoretical bases for favoring one item order over another. Nor, for that matter, is there any compelling reason for using more than one order. It has proved useful, from a practical standpoint, to order the test items so that the apprehensibility of each attribute is tested once with every seventh item, as well as to vary the vowel context such that the eight vowels are cycled every eight items. One ordering yielded by this procedure is indicated by the numbers to the left of the items in Table 3, and it is suggested that this ordering be incorporated as a standard of DRT testing procedure, except where special circumstances may dictate otherwise.

For general testing purposes, the list of test items is cycled four times ("normal administration"), one stimulus word being selected from each item or word pair on each cycle, to yield a total of 448 stimulus words (including 64 experimental words). Depending on the design of the listener's answer sheet, additional "filler items" may be used to absorb the effects of distraction or delay occasioned by page changes, etc. A typical answer sheet is shown in Figure 1. The first item in each column is a filler item, as are the eighth and every seventh item thereafter.

Selection of the stimulus word from each pair can be effectively random in each instance but for the requirement that each

PEST - TEST	FAN - PAN
VAULT - FAULT	CHOCK - JOCK
DUES - NEWS	NOTE - DOTE
VEE - BEE	TICK - THICK
THANK - SANK	CARE - CHAIR
ROD - WAD	DONG - BONG
SO - SHOW	YOU - RUE
LID - RID	REEK - LEAK
DENSE - TENSE	GAFF - CALF
BOSS - MOSS	BOMB - MOM
FOO - POOH	DOUGH - THOUGH
ZEE - THEE	GILT - JILT
FAD - THAD	PENT - TENT
HOP - FOP	YAWL - WALL
ROW - LOW	LOOT - ROOT
GIN - CHIN	VEAL - FEEL
BEND - MEND	NAB - DAB
CHAW - SHAW	BON - VON
JUICE - GOOSE	SOLE - THOLE
PEAK - TEAK	THIN - FIN
BAT - GAT	KEG - PEG
ROCK - LOCK	LONG - WRONG
GOAT - COAT	TUNE - DUNE
MIT - BIT	MEAT - BEAT
THEN - DEN	SHAD - CHAD
GAUZE - JAWS	GOT - JOT
NOON - MOON	DOLE - BOWL
KEY - TEA	DILL - GILL
RAMP - LAMP	LEND - REND

Fig. 1. Specimen DRT Answer Sheet

stimulus word occur twice in the course of the administration and, thus, that each state of each attribute be represented an equal number of times in each vowel context. It is of some advantage to require on occasion that the two halves of a normal administration be at least "balanced," i.e., that each state of each attribute be given equal representations in each vowel context in each half of the test. These constraints serve to partition the test into two identically equivalent halves and grossly equivalent quarters and thus provide some opportunity for evaluating the consistency of the listener's performance during the course of a test.

Recording of Stimulus Materials

For purposes of equipment or system evaluation, the test words are normally recorded without a carrier phrase at a rate of one word per 1.3 - 1.5 seconds. Rates of this order have been found (Cohen)[18] to yield higher scores and smaller standard errors than faster or slower rates, and of course make somewhat more efficient use of testing time than do the rates normally used with various of the more conventional tests of consonant apprehensibility. When the purpose of the test is to evaluate the listener (particularly with very young or handicapped listeners), slower rates of stimulus presentation may be used.

An additional time interval is provided between answer sheets to give listeners ample time to turn from one sheet to the next.

A "filler" item is also recorded at the place corresponding to the top of each column on the listener's answer sheet to provide further insulation against any distraction that might be occasioned by spatial disparities between successive items on the listener's answer sheet.

No attempt is made to achieve a uniform level from one test word to the next, but an attempt is made to establish a fixed recording level which will yield an average vowel peak value of -2 VU. On completion of the editing process, averaged vowel peak values are then used as a basis for setting the level of a 1 KHz calibration tone which is recorded at the beginning of each tape.

Speakers normally require some amount of practice to achieve uniform, rhythmic delivery in synchrony with a timing light. They are instructed only to "speak in a normal, conversational manner -- avoid over-enunciation." The rhyming option of each stimulus word is shown next to the stimulus word on the speaker's script in order to minimize ambiguity in pronunciation. Subject to the results of research in progress, it may prove feasible to coach the speaker in various ways to achieve a more "normal" manner of enunciation as defined by his "diagnostic profile" under various transmission conditions.

Selection of Speakers

The problem of speaker selection for purposes of evaluating equipment or listener characteristics is yet to find a generally satisfactory solution. The hazards associated with arbitrary selection of single speakers are evident from the literature. It is unlikely, however, that the use of two or three haphazardly selected speakers is sufficient to assure the generality of results, whereas practical considerations often preclude the use of substantially larger numbers of speakers. Until all of the relevant speaker variables have been identified, the problem of speaker selection can be dealt with only in a tentative and, necessarily, somewhat arbitrary manner.

In one attempt to devise a means of selecting a "typical voice," a semantic differential-type voice rating form was used to select from a pool of 32 speakers one voice which was judged most nearly neutral with respect to a set of four <u>perceived voice traits</u> (PVT's) as described by Voiers.[19]

Subsequently, it has appeared that the DRT itself is sensitive in a number of dimensions to differences among speakers and may thus provide an effective means of selecting speakers of desired characteristics.

<u>Selection and Training of Listeners</u>

A crew of eight, minimally trained listeners has been found sufficient for most purposes of equipment evaluation with the DRT, although a smaller crew may suffice, depending on the level of precision desired. Crews of eight listeners typically yield standard errors on the order of 1% (adjusted for chance) over most of the range of possible scores. However, slightly larger values obtain toward the lower end of the intelligibility scale. Because the text exhibits a degree of listener sensitivity, however, care should be exercised in selecting listeners for tests conducted to evaluate speakers or communications equipment. Clinically normal hearing below 6,000 Hz is desirable. Standards based on performance on the DRT itself have been found useful for purposes of equipment evaluation.

Administration of the Test

The use of "live" test presentation procedures tends to be somewhat impractical for most purposes, and the use of pre-recorded materials, as described above, is thus to be preferred in general. For routine purposes of system evaluation, an average vowel peak level of approximately 72 dB SPL (flat plate) appears to be most satisfactory.

Listeners are instructed simply to strike out the member of each word pair that they perceive to be the stimulus word. It is stressed that there are no "right answers" other than those dictated by the listener's perceptions of the stimulus words.

Scoring the Diagnostic Rhyme Test

DRT response data can be scored in a diversity of ways, depending upon the interests of the investigator. Generally, however, greatest interest will attach to the six major "diagnostic" scores, each constituting an indicant of the gross apprehensibility of the speaker's intent with respect to a given attribute. It is possible, in addition, to fractionate each of the major diagnostic scores into various components (e.g., to obtain separate scores for the apprehensibility of _sustention_ in the voiced and unvoiced planes; _voicing_ in the frictional - non-frictional planes; _nasality_ in different vowel contexts, and so on).

Separate scores for the apprehensibility of each state of each attribute are likely to be of interest in that some experimental variables may affect the apprehensibility of the two states of some attributes in an asymmetrical manner. The resulting discrepancy between listener scores for the two states of an attribute is termed _bias_. It is measured simply as the difference between the percent (adjusted for chance) of the time listeners correctly apprehend the positive state (e.g., voiced) of an attribute and the percent of the time they correctly apprehend the negative state (e.g., unvoiced).

Finally, a total score, representing the average of the six major diagnostic scores is likely to be of interest in many applications. Research with previous versions of the DRT has shown that such scores are generally equivalent, numerically, to scores yielded by the Fairbanks Rhyme Test, but there is some indication (Voiers _et al._)[20] that the DRT is sensitive to certain types of deficiencies not reflected in FRT scores.

In principle, at least, DRT results lend themselves to expression in terms of signal detection theory or information theory. However, a somewhat simpler approach to the scoring problem provides a solution which is probably adequate for most practical purposes and also most consistent with prevailing conventions. It involves the familiar correction for guessing, accomplished by means of the following formula:

$$S = \frac{100\ (R - W)}{(T)}$$

where S is the "true" percent-correct responses, R is the observed number of correct responses, W is the observed number of incorrect responses, and T is the total number of items involved. <u>This correction is applied to all DRT scores, including the gross or total score</u>.

Manual scoring of the DRT through the use of templates is quite feasible where the investigator is concerned only with obtaining a gross score and perhaps the six major diagnostic scores. However, computer scoring not only facilitates this process, particularly where multiple scramblings of the test materials are involved, but also provides easy access to a wealth of other potentially useful data. Among these are separate tallies of individual listener errors in the apprehension of each state of each attribute; error counts for individual items; and total errors per subject for various subdivisions of the test. This last serves, in light of the systematic redundancy of the DRT, to provide a powerful check on the state of alertness of individual listeners over the course of the test and for keypunching errors during the transcription of test data for computer analysis. A specimen printout for one scoring scheme is shown in Fig. 2.

Validity of the Diagnostic Rhyme Test

It is not possible within the scope of this report to treat all aspects of the issue of the validity of the DRT, but it is appropriate at least to address the major issue regarding the validity of the DRT and the concepts on which it is based. Obviously, the value of the DRT would be greatly restricted if it proved insensitive to qualitative differences in the effects of different forms of speech signal impoverishment. It has, in fact, proven highly sensitive to such differences and yielded results consistent with known facts of acoustic phonemics. Fig. 2 thus serves to illustrate the diversity of diagnostic patterns yielded with some common forms of speech degradation. Represented in the figure are speech high passed at 4 KHz, low passed at 800 Hz, and noise masked with a S/N ratio of + 3dB. Also represented are the averaged scores for a sample of present-day digital vocoders.[20]

All data represents averages for two administrations of the DRT for each of six male speakers. A crew of eight male listeners was used. There are important similarities and differences among the results for the four conditions. They show, for one

AVERAGE BY SPEAKERS ACROSS LISTENERS FOR 'C' TAPES 10 MAY 72 PAGE 2

	PRESNT	S.E.	ABSENT	S.E.	BIAS	S.E.	TOTAL	S.E.
VOICING	97.2	.83	96.7	1.05	.5	1.13	97.0	.76
FRICTIONAL	95.6	1.57	94.7	1.80	.9	2.42	95.1	1.18
NONFRICTIONAL	98.8	.72	98.8	.44	.0	.61	98.8	.51
NASALITY	99.2	.29	99.5	.13	-.3	.34	99.3	.14
GRAVE	99.0	.44	99.1	.31	-.1	.55	99.0	.26
ACUTE	99.5	.16	99.9	.13	-.4	.17	99.7	.12
SUSTENTION	98.0	.27	98.4	.65	-.3	.60	98.2	.40
VOICED	97.9	.77	98.4	.29	-.5	.69	98.2	.47
UNVOICED	98.2	.39	98.3	1.09	-.1	1.11	98.2	.60
SIBILATION	98.9	.51	99.5	.13	-.6	.54	99.2	.25
VOICED	98.3	1.09	99.2	.20	-.9	1.13	98.8	.55
UNVOICED	99.5	.16	99.7	.16	-.3	.16	99.6	.14
GRAVENESS	94.3	1.39	97.1	.95	-2.8	2.01	95.7	.64
VOICED	98.7	.39	99.7	.16	-1.0	.48	99.2	.17
UNVOICED	89.8	2.73	94.4	1.89	-4.6	3.95	92.1	1.27
STOPPED	99.0	.52	99.2	.64	-.3	1.00	99.1	.30
UNSTOPPED	89.6	2.48	94.9	1.61	-5.3	3.50	92.3	1.14
COMPACTNESS	99.6	.17	99.5	.16	.1	.22	99.5	.13
VOICED	99.9	.13	99.5	.16	.4	.17	99.7	.12
UNVOICED	99.3	.37	99.5	.26	-.1	.42	99.4	.24
SUSTAINED	99.5	.26	99.5	.33	.0	.45	99.5	.19
INTERRUPTED	99.7	.16	99.5	.16	.3	.26	99.6	.10
B/M	99.9	.13	99.2	.29	.7	.31	99.5	.16
B/E	99.3	.24	99.7	.16	-.4	.27	99.5	.16
EXPERIMENTAL.	99.5	.12	99.7	.26	.1	.22	99.6	.17

SPKR(S)=	RD	JE	CH	BV	SN	BL
LIST(S)=	104A	111A	107A	103B	115B	116B

NUMBER OF SPEAKERS 6
WORDS PER SPEAKERS 4608

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
X TOTAL DRT SCORE= 98.2 X
X STANDARD ERROR= .17 X
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

Fig. 2. Specimen Report of Diagnostic Rhyme Test Results.

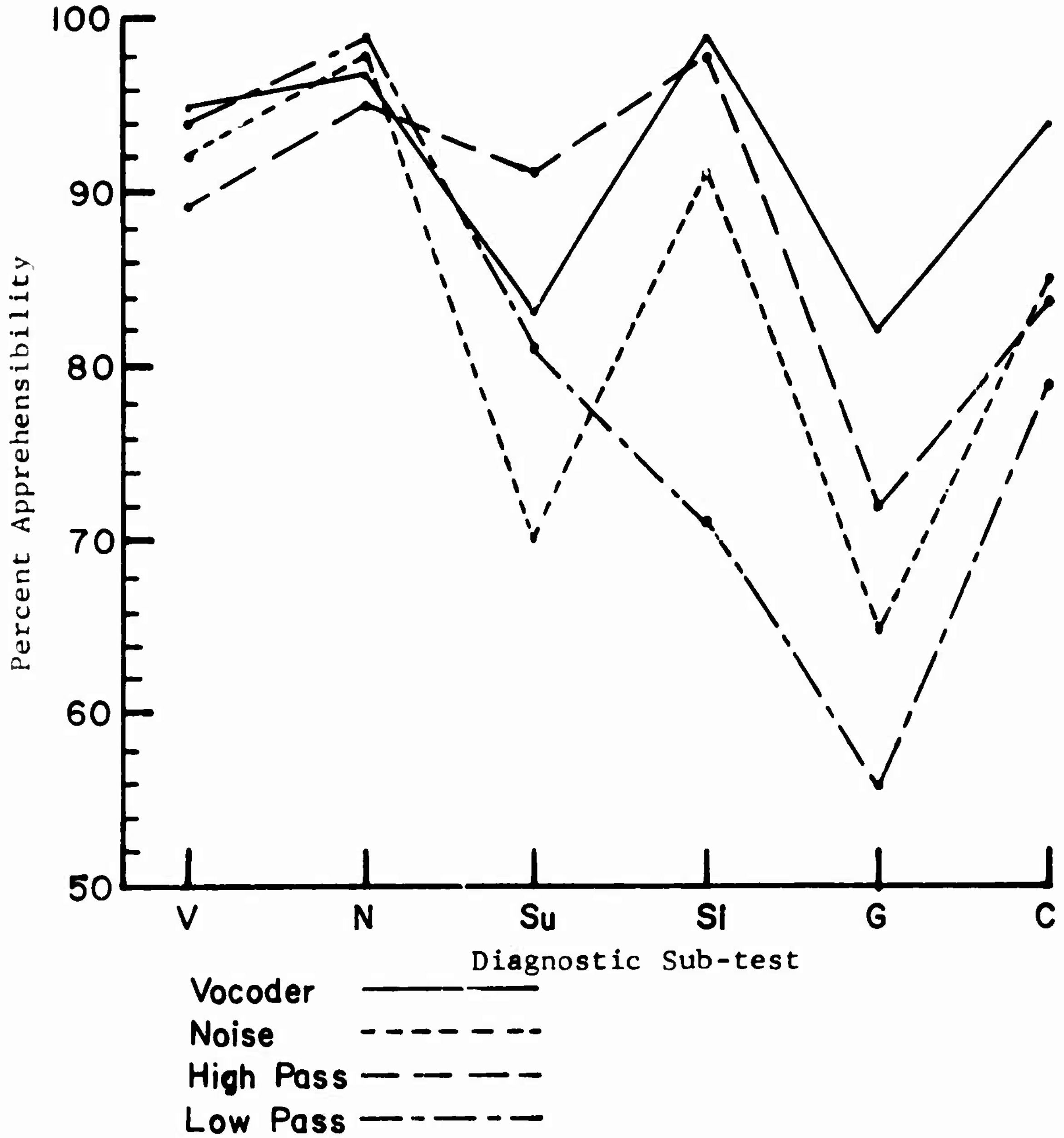

Fig. 3. Diagnostic Patterns for Four Transmission Conditions.

thing, that the sustained-interrupted and grave-acute distinctions tend rather generally to be most difficult and most susceptible to speech impoverishment. Voicing and nasality, on the other hand, retain a high level of apprehensibility under most conditions of signal impoverishment. Voicing does not, however, remain equally apprehensible under all conditions and is predictably, perhaps, relatively more apprehensible under low pass than high pass conditions.

Particular interest possibly attaches to the comparison of results for noise masked and low passed speech. As many investigators have noted, band limited Gaussian noise has the effect of high frequency attenuation due to the relatively low level of speech energy in the higher frequencies of the speech spectrum. The diagnostic patterns found here to characterize the two cases are in fact quite similar in most respects. They are readily differentiated, however, on the basis of the sibilation scale of the DRT. Predictably, low pass filtering greatly reduces the apprehensibility of sibilation. Less predictably, however, noise has relatively little impact upon the apprehensibility of this attribute, in spite of the fact that noise is itself the major acoustical correlate of the attribute. Differences between the diagnostic patterns for high passed and low passed speech are of a generally predictable character. Their similarities in terms of graveness are also predictable in that the ranges of the second and third formants were largely excluded by both pass bands.

The foregoing results attest to one aspect of the validity of the DRT, its sensitivity to qualitation differences in the characteristics of transmission channels or media. Various other aspects of this issue will be dealt with in forthcoming reports.

REFERENCES

[1]G. A. Miller, G. A. Heise and W. Lichten, "The Intelligibility of Speech as a Function of the Context of the Test Materials," J. Exper. Psychol., 41, 329-335 (1951).

[2]D. Howes, "On the Relation Between the Intelligibility and Frequency of Occurrence of English Words," J. Acoust. Soc. Amer., 29, 296-305 (1957).

[3]D. J. Bruce, "The Effect of Listener's Anticipations on the Intelligibility of Heard Speech," Language and Speech, 1, 79-97 (1958).

[4]R. W. Peters, "The Effect of Length of Exposure to Speaker's Voice on Listener Reception," Joint Project Report No. 44, The Ohio State University Research Foundation and U. S. Naval School of Aviation Medicine (1955).

[5]Miller *et al.*, *op. cit.*

[6]J. P. Egan, "Articulation Testing Methods," Laryngoscope, 58, 955-991 (1948).

[7]G. A. Miller and P. Nicely, "An Analysis of Perceptual Confusions Among Some English Consonants," J. Acoust. Soc. Amer., 27, 338-352 (1955).

[8]G. Fairbanks, "Test of Phonemic Differentiation: The Rhyme Test," J. Acoust. Soc. Amer., 30, 596-600 (1958).

[9]A. S. House, C. E. Williams, H. L. Hecker, and K. D. Kryter, "Articulation Testing Methods: Consonantal Differentiation with a Closed Response Set," J. Acoust. Soc. Amer., 37, 158-166 (1965).

[10]F. R. Clarke, "Technique for Evaluation of Speech Systems," Final Report, Contract No. DA28-043-AMC-00 227 (E) USAEL (1965).

[11]R. Jakobson, C. G. M. Fant and M. Halle, "Preliminaries to Speech Analysis: The Distinctive Features and Their Correlates," Tech. Rep. No. 13, Acoustics Laboratory, MIT (1952).

[12]Miller and Nicely, *op. cit.*

[13]P. Delattre, "From Acoustic Cues to Distinctive Features," Phonetica, 18, 198-230 (1968).

[14]Miller and Nicely, op. cit.; Jakobson, Fant and Halle, op. cit.; and Delattre, op. cit.

[15]I. Pollack, H. Rubenstein, and C. Decker, "Intelligibility of Known and Unknown Message Sets," J. Acoust. Soc. Amer., 31, 273-279 (1959).

[16]M. Halle, "On the Bases of Phonology," in J. A. Fodor and J. J. Katz, The Structure of Language, Prentice-Hall, Inc., Englewood Cliffs, N. J. (1964).

[17]Miller and Nicely, op. cit.

[18]M. F. Cohen, "Effects of Stimulus Presentation Rate Upon Intelligibility-Test Scores," J. Acoust. Soc. Amer., 37, 1206 (1965), (A).

[19]W. D. Voiers, "Perceptual Bases of Speaker Identity," J. Acoust. Soc. Amer., 36, 1065-1073 (1964).

[20]W. D. Voiers, "Performance Evaluation of Speech Processing Devices, III. Diagnostic Evaluation of Speech Intelligibility," Final Report, Contract No. AF19(628)4987, AFCRL (1967).

[21]W. D. Voiers and C. Smith, "Diagnostic Evaluation of Intelligibility in Present-Day Digital Vocoders," AFCRL-72-0120, 22 February, 1972, Special Reports, No. 131, 170-175 (IEEE Cat. No. 72) CHO 596-7 AE.

CHAPTER 2

THE NATURE OF INDIVIDUAL DIFFERENCES IN DIAGNOSTIC RHYME TEST PERFORMANCE

by

William D. Voiers and Alan D. Sharpley

INTRODUCTION

Sampling error associated with listeners is a perennial problem for the investigator who uses the response of human listeners to evaluate the performance of speech communication and processing equipment. The precision of such evaluations, and in turn the power of statistical tests performed in conjunction with them, varies inversely with degree of inter-listener variation. Thus methods of controlling inter-listener variation, whether by statistical or experimental means, offer possibilities for enhancing the precision or reliability of intelligibility test results. However, the development of such methods presupposes some understanding of the origin and nature, as well as the degree. of inter-individual variation in speech discrimination ability.

The effects upon speech perception of individual differences associated with pathology have been extensively investigated. Generally, major emphasis has been upon the degree rather than the nature of the discriminative deficiencies associated with various pathological conditions. In one case, however, the Diagnostic Rhyme Test was used to investigate the effects of pathology upon specific speech discrimination abilities, and provided some valuable insights.[1] This investigation revealed quite clearly that the effects of sensori-neural hearing impairment on speech discrimination performance are of a highly specific rather than general character. Depending upon degree and nature of hearing impairment, different aspects of speech discrimination performance are affected. Within the clinical population, at least, speech discrimination ability is not a unidimensional entity.

The nature of inter-individual differences in speech discrimination ability in the normal hearing population is yet to be extensively investigated. However, the results of an investigation by Elliott _et al_.[9] throws some light on the issue. These investigators employed factor analytic techniques in an attempt to identify the correlates of verbal recognition ability as measured by the Fairbanks Rhyme Test. They found performance on the Fairbanks test to be correlated with individual differences on both auditory and "non-auditory" tests. Among the "non-auditory" correlates of Rhyme Test performance were: _vocabulary test performance_, _word fluency_. In both cases correlations with performance on the Fiarbanks Rhyme Test were positive. Auditory discrimination measures that correlated significantly with Fairbanks Rhyme Test performance were _absolute thresholds for pure tones_ and _difference thresholds for tonal duration_, _frequency and intensity_. Unexpectedly, however, the correlations between Rhyme Test scores and absolute threshold measures were _negative_, which fact implies that hearing loss (at least over the range involved) is associated with superior performance on the Fairbanks Rhyme

Test. Of the seven factors revealed by the factor analysis, the Fairbanks Rhyme Test exhibited substantial loadings on five, including a factor defined primarily by measures of intellectual aptitude.

Given that speech discrimination ability as measured by the Fairbanks Rhyme Test has such a diversity of antecedents, the question arises as to whether speech discrimination involves a single ability or a number of independent abilities. Is it in fact a single, global ability, or a congeries of more elementary abilities. Because of the diversity of measures it yields, the Diagnostic Rhyme Test is eminently adapted to the purpose of resolving this issue. Accordingly, a factor analytic investigation of individual differences in Diagnostic Rhyme Test performance was undertaken.

METHOD AND MATERIALS

Subjects

Subjects for this investigation were 72 male college students from the University of Texas, all of whom were born and raised in the United States. Their ages range from 17 - 36. They were paid at the rate of $2.00/hr. to participate in this and related investigations.

Speaker

A single, male speaker (RD) recorded all of the speech materials in this investigation. He was selected on the basis of research results which revealed him to have a highly typical DRT diagnostic score pattern under a diversity of transmission conditions.

Test Materials

Subjects were administered the following tests in random groups of eight:

1. Diagnostic Rhyme Test III (nine administrations, different randomizations, the first two of which yielded data used in this investigation).

2. Fairbanks Rhyme Test (five administrations, different randomizations, the first two of which yielded data used in this investigation).

3. Cooperative English Test - Form 1B - I (A four-choice

test of English vocabulary).

4. Cooperative English Test - Form 1A - I (A four-choice test for the effectiveness of English expression).

5. Wide Range Vocabulary Test (A five-choice test of English vocabulary).

6. Word Productiveness Test (A test of the ability to produce words with common initial consonants -- j, g, b, h).

7. Pure Tone Audiometric Tests (Two administrations, Rudmose ARJ-4A Békésy recording audiometer; audiometric data for each subject's "best ear" were used in the analysis. The "best ear" was selected on the basis of lower total loss across the five frequencies tested).

8. Minnesota Multiphasic Personality Inventory (A series of preliminary analyses failed to reveal any significant personality correlates of speech discrimination performance. Accordingly, data from this test are not treated in the present investigation).

Scores for each subject on 17 variables were obtained with the test materials described above. Data on these variables were then used for purposes of a factor analytic examination of individual differences in speech discrimination. The variables treated in the analysis were:

1. DRT - Total Diagnostic Rhyme Test (DRT) percentage score*

2. VOIC - Score on the Voicing sub-test of the DRT*

3. NASL - Score on the Nasality sub-test of the DRT*
4. SUST - Score on the Sustention sub-test of the DRT*
5. SIBI - Score on the Sibilation sub-test of the DRT*
6. GRAV - Score on the Graveness sub-test of the DRT*
7. CMPT - Score on the Compactness sub-test of the DRT*
8. FRT - Fairbanks Rhyme Test percentage score*
9. VOCB - Cooperative English Test (vocabulary) percentage score
10. EFCT - Cooperative English Test (effectiveness) percentage score
11. WRVT - Wide Range Vocabulary Test percentage score
12. WPT - Word Productiveness Test - average number of words produced for four initial consonants
13. 1K - Hearing loss (dB re ISO-1964 standards) at 1000 Hz*
14. 2K - Hearing loss at 2000 Hz*
15. 3K - Hearing loss at 3000 Hz*
16. 4K - Hearing loss at 4000 Hz*
17. 6K - Hearing loss at 6000 Hz*

*Average score for two administrations

RESULTS AND DISCUSSION

The matrix of product-moment correlations among the seventeen variables under investigation is presented in Table 4. Coefficients of reliability are shown in the cells of the major diagonal axis. Several aspects of these results merit comment, for example, the correlation between FRT and total DRT score which, though positive, is negligible. Evidently the two tests tap somewhat different aspects of speech discrimination ability, and only VOIC and SIBL exhibit significant ($p<.01$) correlation with the FRT. Negligible correlations obtain for the cases of all other DRT sub-tests.

It is also noteworthy that no measure of speech discrimination ability exhibits a significant positive correlation with any measure of auditory sensitivity. In fact, the only correlations which approach statistical significance are of negative sign. However, in contrast with the results of Elliott _et al_., all correlations between measures of auditory sensitivity and FRT performance are in the positive direction, though of negligible magnitude.

Various other aspects of Table 4 would merit discussion, but the issues on which they bear are brought into somewhat clearer focus by means of factor analysis. Factor analysis of the correlation matrix in Table 4 yielded seven orthogonal factors which accounted for 93 percent of the systematic variations among

TABLE 4 Correlation Matrix for Seventeen Listener Variables

	DRT	VOIC	NASL	SUST	SIBL	GRAV	CMPT	FRT	VOCB	EFCT	WRVT	WPT	1K	2K	3K	4K	6K	MEANS	S.D.
DRT	.54																	98.79	.88
VOIC	.50	.41																98.22	2.07
NASL	.43	.16	.00															99.20	1.13
SUST	.59	-.03	-.03	.69														98.00	3.04
SIBL	.49	.24	.28	-.06	.48													99.00	1.71
GRAV	.50	.06	.32	.13	.10	.48												99.02	1.48
CMPT	.29	-.03	-.01	.08	.06	.10	.35											99.31	1.16
FRT	.13	.27	-.15	-.03	.26	.01	-.11	.64										99.02	.89
VOCB	.17	.24	.00	.00	.02	.12	.13	.16	.95									77.08	12.24
EFCT	.04	.16	.00	.03	-.17	.11	-.05	.07	.76	.95								75.56	13.36
WRVT	.36	.23	.05	.19	.24	.15	.16	.32	.74	.55	.95							80.79	6.28
WPT	.37	.27	.10	.18	.27	.10	.11	.01	.34	.26	.36	.87						23.47	4.59
1K	.00	-.07	.05	.00	-.07	.07	.11	.14	.10	.04	.03	-.02	.94					-2.42	4.54
2K	-.10	-.24	-.07	.00	-.07	.05	.09	.05	-.07	-.13	-.11	-.08	.75	.95				-2.89	5.75
3K	-.06	-.12	.06	-.06	.06	-.01	-.03	.06	.03	-.03	.00	-.09	.57	.62	.97			-6.74	8.31
4K	-.20	-.16	.02	-.12	-.09	-.08	-.11	.07	.00	-.04	-.06	-.05	.59	.58	.73	.91		-5.79	7.06
6K	.00	-.20	.00	.05	.17	.03	-.03	.05	-.06	-.04	-.03	.02	.31	.36	.45	.47	.96	-14.18	14.86

For 70 df: $r = .23$, $p < .05$; $r = .30$, $p < .01$

listeners. Rotation of axis to a varimax criterion of simple structure yielded the pattern of loadings shown in Table 5.

Factor I is defined by the various measures of verbal aptitude. No other variables have significant loadings on this factor.

Factor II is defined by measures of auditory sensitivity. No other variables have significant loadings on this factor. Although Elliott _et al_. observed negative correlations between measures of auditory sensitivity and Fairbanks Rhyme Test scores, no such relation is indicated here. Nor is there any indication that DRT performance depends to any degree on auditory sensitivity to pure tone stimuli, at least within the range of auditory sensitivity characteristic of this sample of listeners. As noted earlier, however, the DRT is sensitive to auditory deficiencies of pathological magnitude.

Factor III is defined primarily by the DRT sub-test for the apprehensibility of _sustention_. Several other variables have appreciable loadings on this factor, but the FRT would appear to be insensitive to this dimension of inter-individual variation.

Several variables contribute to the definition of Factor IV. They include the Fairbanks Rhyme Test, the Wide Range Vocabulary Test scores, and scores for the two DRT attributes, _voicing_ and _sibilation_. The loading of WRVT indicates a positive relationship between verbal ability (as measured by vocabulary) and

TABLE 5. Factorial Structure of Seventeen Listener Variables

	I	II	III	IV	V	VI	VII
DRT	.06	-.08	.49	.29	-.02	.25	.63
VOIC	.15	-.15	-.05	.40	-.17	.22	.29
NASL	-.03	.01	-.02	.00	.01	.08	.49
SUST	.04	-.08	.81	-.01	.12	.02	.06
SIBL	-.12	-.05	-.04	.47	.16	.30	.39
GRAV	.11	.03	.21	-.04	-.04	-.07	.61
CMPT	.03	.08	.27	-.06	-.17	.14	.13
FRT	.13	.10	.00	.75	.01	-.09	-.12
VOCB	.90	.05	.00	.11	-.09	.16	.07
EFCT	.92	-.05	-.03	-.11	.01	.01	.02
WRVT	.74	-.02	.24	.40	.01	.17	.10
WPT	.25	-.05	.15	.05	.02	.85	.09
1K	.07	.89	.11	.03	-.10	-.02	.03
2K	-.11	.88	.15	-.06	-.01	-.04	-.08
3K	.02	.79	-.15	.03	.32	-.06	.11
4K	.01	.77	-.20	-.03	.32	.00	-.08
6K	-.04	.37	.06	.01	.85	.03	.03

scores on the FRT, which finding is generally consistent with the results of Elliott _et al_. The loading of VOIC is in line with previous observations concerning the structure of the FRT, but the SIBL loading was somewhat surprising, since one possible deficiency of the FRT is the negligible demand it makes upon the listener with respect to this attribute of consonant phonemes.[4]

WPT defines Factor VI, and several DRT variables exhibit substantial loadings on this factor. The FRT, however, has a negligible loading. Possibly this factor relates to some aspect of perceptual motor speed or test-taking skill. The rapid pace at which listeners must work in taking the DRT (one response every 1.4 seconds) might thus account for the loadings exhibited by various DRT variables. A question arises, however, as to why the FRT, which involves the same stimulus presentation rate, does not exhibit a high loading. The answer to this question is not clear, but one possibility derives from the fact that all listeners were given extensive exposure to the DRT before taking the FRT. Possibly, therefore, they were more nearly habituated to the time pressures involved by the time they took the FRT.

Factor VII evidently represents a dimension of speech discrimination skill in that the total DRT score and three of its components -- nasality, sibilation and graveness -- have substantial loadings on this factor. The slightly negative loading for FRT is somewhat puzzling, but can probably be attributed to chance.

From the foregoing it is evident that at least three independent factors (III, IV and VII) contribute to listener variation in speech discrimination performance. The first of these (defined by SUST) appears to be related to the ability to discriminate characteristics of the speech envelope while the second appears to involve the ability to detect the presence and character of noise. The third appears to involve the ability to discriminate the characteristics and relationship of the first three formants. Other dimensions might have emerged but for the fact that all speech materials were presented under high fidelity conditions, which circumstance may have operated to minimize inter-listener variation in potentially significant dimensions of discriminative ability.

It appears that the FRT is a relatively unitary measure, loading substantially on only a single factor. Conceivably, therefore, it fails to tap certain aspects of the speech discrimination task. The DRT, on the other hand, has fairly high loadings on all factors involving speech discrimination, and would thus appear to provide a more comprehensive measure of the adequacy of a listener's discriminative capacity.

Clearly, additional research will be required to resolve the issue completely, but the results of this investigation strongly suggest that speech perception involves more than one dimension of inter-individual variation in discriminative capacity. This suggestion has obvious implications for the development of procedures for selecting operational communication personnel as well as listeners to be used in the research and testing situations.

REFERENCES

[1] M. H. Olroyd, "Employment of the Diagnostic Rhyme Test (DRT) with Normal-Hearing and Sensori-Neural Hearing-Impaired Listeners," unpublished dissertation, Louisiana State University (1972).

[2] D. N. Elliott, W. D. Riach, J. P. Sheposh, and C. Trahiotis, "Discrimination Performance of High School Sophomores on a Battery of Auditory Tests," Acta Oto-Larynglogica, Supplementum 216 (1966).

[3] Olroyd, op. cit.

[4] W. D. Voiers, "Performance Evaluation of Speech Processing Devices, III. Diagnostic Evaluation of Speech Intelligibility," Final Report, Contract No. AF19 (628) 4987, AFCRL (1967).

CHAPTER 3

SPEAKER EFFECTS ON INTELLIGIBILITY TEST RESULTS

by

William D. Voiers and Carl J. Hehmsoth

SPEAKER EFFECTS ON INTELLIGIBILITY TEST RESULTS

THE PROBLEM

The possible effects of a speaker's idiosyncracies upon the results of intelligibility tests conducted to evaluate communications equipment has long been a matter of concern to investigators in the field of speech communication. But while it is clear that speaker effects exist, the nature of these effects has not been extensively investigated.

The distinction between *general* effects and *interactive* effects is particularly important in this context. To the extent that differences among speakers tend to remain constant across transmission conditions or situations, the speaker effect involves a *general* component. To the extent that speaker differences vary from one transmission condition to the next, the speaker effect involves an *interactive* component.

The obvious practical consequence of any type of speaker effect is that, normally, systems evaluated with one speaker cannot be directly compared to systems evaluated with a different speaker. However, the possibility may exist of independently evaluating *general* differences among speakers and in turn adjusting results obtained with individual speakers in such a way as to render them comparable.

To the extent that speaker idiosyncracies are _interactive_ with transmission conditions (i.e., to the extent that different systems may respond to different voices in different ways) system comparisons involving different speakers are potentially invalid. Control of such effects is, moreover, difficult to accomplish by means other than those involving the use of large samples of speakers.

In addition to the effects of gross differences in speaker intelligibility, general and interactive, there exists the possibility that speakers differ systematically in terms of the apprehensibility of specific speech features. Thus speakers who yield comparable measures of gross intelligibility under a given condition may nevertheless be characterized by qualitative differences in intelligibility, i.e., the discriminability of certain speech features may vary from one speaker to the next. Such effects have obvious implications for the technology of _diagnostic_ intelligibility testing.

A comprehensive treatment of the issues raised here is beyond the scope of the present project. However, various results obtained in the course of the project provide some insights regarding them. The results of two experiments, in particular, are relevant in this connection.

EXPERIMENT I

Methods and Materials

Speakers. Twelve male speakers, selected primarily on the basis of availability, were used in this investigation. Their ages ranged from 20 to 45.

Listening Crew. The listening crew was composed of eight males between the ages of 18 and 24. All members of the crew had extensive experience with the Diagnostic Rhyme Test.

Test Materials. The Diagnostic Rhyme Test (Form III) was used for purposes of this investigation.

Test Conditions. Diagnostic Rhyme Test materials as recorded by each of the twelve speakers were presented to the listening crew under a diversity of transmission conditions. Five of these were selected for the illustrative purposes of this investigation. They included:

1. Undegraded speech
2. Low passed (400 Hz) speech
3. High passed (3 KHz) speech
4. Noise masked (-10dB S/N)
5. Digitally vocoded (1200 bps)

The level of the speech signal, prior to processing, was approximately 72 dB SPL in the first four conditions. The vocoded speech was presented to listeners at this same level. All tests were conducted in partitioned IAC rooms. The test materials were presented diotically over TDH-39 earphones mounted in Rudmose Otocups.

Results

The analyses of results reported here are addressed to the following issues:

1. Consistency across conditions of speaker order with respect to gross intelligibility.
2. Consistency across test conditions of diagnostic score patterns of individual speakers.

DRT total scores for the twelve speakers were ranked for each of the five test conditions. The results are presented in Table 6.

TABLE 6. Ranked DRT Scores of Twelve Speakers Under Five Transmission Conditions

	Transmission Conditions				
Speaker	Undegraded	High-Pass	Low-Pass	Noise	Vocoded
A	8	11	6	7	7
B	12	1	12	8	12
C	6	8	5	6	6
D	9	12	3	11	3
E	3	10	4	4	9
F	4	3	2	1	8
G	10	5	10	10	10
H	11	7	7	9	4
I	7	6	8	5	5
J	2	4	9	3	2
K	1	2	1	2	1
L	5	9	11	12	11

It is evident from the table that, while speaker ranks under the various conditions are by no means perfectly intercorrelated, a high degree of intercorrelation exists. Generally, speakers who rank high under one condition tend to maintain similar ranks under other conditions. The most notable exception occurs in the case of speaker B. Ranked below average on all other conditions, he achieves the top rank in the case of high passed speech. The reasons for this inversion are not evident. The fact that speaker B's voice is the highest pitched in this sample is of possible interest. There is, however, no indication otherwise that high-pitched voices are more intelligible under high pass conditions.

Table 7 presents the correlations among speaker ranks for the five test conditions involved here.

TABLE 7. Correlations (ρ) Among Ranked DRT Total Scores of Twelve Speakers for Five Transmission Conditions.

Condition	Undegraded	High-Pass	Low-Pass	Noise	Vocoded
Undegraded	--				
High-Pass	.10	--			
Low-Pass	.48	-.16	--		
Noise	.68	.47	.53	--	
Vocoded	.37	-.02	.54	.36	--

It is evident from the table that speaker ranks are not equally predictable from any one condition to another, although the size of the sample involved here permits only the most tentative conclusions.

Clearly, intelligibility measures obtained under conditions involving high-passed speech are of little value in predicting a speaker's relative level of intelligibility under other transmission conditions. However, the level of predictability among the various other conditions examined here is at least more than negligible in all instances and relatively high in several. In particular, a speaker's relative intelligibility under high fidelity conditions correlates quite well (.68) with his relative level under noisy conditions. More generally, however, it must be concluded that speaker characteristics are interactive with channel characteristics and thus that the results of system comparisons involving a single speaker may be of questionable validity.

In addition to the issue of gross quantitative differences among speakers, there is also the issue of qualitative differences. To what extent do speakers differ, for example, in terms of diagnostic score patterns? Are such differences general in nature or interactive with transmission conditions? To throw some light on this issue, diagnostic scores yielded by the twelve speakers discussed above were examined under the same five transmission conditions.

For this purpose diagnostic data for each transmission condition were adjusted to remove the effects of speaker differences in total DRT score. Data so adjusted were then analyzed to obtain for each speaker an average deviation score, i.e., the average of the absolute differences between his adjusted scores on DRT sub-tests and the average score of the group on corresponding sub-tests. The average so obtained thus represented for each speaker an indicant of conformity (or nonconformity) with the group under a given transmission condition.

The question then arises as to what extent speakers who yield deviant patterns under one condition tend also to yield deviant patterns under others. Table 8 presents results which bear upon this question.

TABLE 8. Ranked Deviation Scores of Twelve Speakers and Five Transmission Conditions.[1]

Speaker	Transmission Conditions				
	Undegraded	High-Pass	Low-Pass	Noise	Vocoded
A	11	12	6	11	9
B	1	2	4	12	1
C	3	9	11	8	12
D	7	1	3	1.5	6
E	4	4	5	3	5
F	5	3	2	4.5	10
G	10	6	7	6	7
H	8	10	9	10	4
I	9	5	10	9	11
J	12	11	12	7	8
K	2	8	1	4.5	3
L	6	7	8	1.5	2

[1] Higher ranks denote smaller deviation scores, i.e., a rank of "1" identifies the speaker with the most typical diagnostic pattern under a given transmission condition.

From the table it is evident that speakers who yield typical diagnostic score patterns under one test condition tend rather strongly to yield typical patterns under other conditions, but pronounced exceptions to this tendency are evident. Speaker B, for example, yields highly typical diagnostic score patterns under four conditions. In the case of noise-masked speech, however, his pattern is the most deviant of the group.

Table 9 shows the correlations among ranked deviation scores for the five tranmission conditions.

TABLE 9. Correlations ( ) Among Pattern Deviation Scores of Twelve Speakers for Five Transmission Conditions

Condition	Undegraded	High-Pass	Low-Pass	Noise	Vocoded
Undegraded	--				
High-Pass	.47	--			
Low-Pass	.51	.55	--		
Noise	.13	.37	.32	--	
Vocoded	.39	.21	.39	.16	--

In general, the values of the coefficients of correlation are somewhat higher than those in Table 7, suggesting that individual speakers tend more strongly to maintain their diagnostic score patterns from one condition to the next than to maintain their relative level of gross intelligibility. Correlations among speaker ranks are far from perfect, however.

It is clear, therefore, that speaker differences in overall intelligibility and in diagnostic score patterns are interactive with channel or transmission conditions, and that comparative test results obtained with a single speaker may not be generalized with a high degree of confidence to the population of speakers at large. It should be stressed, however, that the present investigation involved comparisons among extremely diverse types of transmission conditions. Such diversity is unlikely to be encountered in practical testing situations. Rather, the systems or transmission conditions typically subjected to comparative evaluation are likely to involve relatively similar types and degrees of speech degradation. The question arises, therefore, as to the practical implications of speaker x system interaction. The scope of the present effort does not permit a comprehensive investigation of this issue, but data obtained in the course of the project throw some light on the issue. They are presented and discussed in the following investigation.

EXPERIMENT II

Methods and Materials

Speakers. Six male speakers, selected on the basis of availability, dialectal characterstics, or pitch frequency were used in this investigation. Two of the six speakers were judged by a listening crew to have voices of higher than average pitch, while two were judged to have voices of lower than average pitch, and two were judged to have voices of average pitch for male speakers.

Listening Crew. The listening crew was composed of eight males between the ages of 18 and 24. All members of the crew had extensive experience with the Diagnostic Rhyme Test.

Test Materials. Recordings of DRT IV were used for purposes of this investigation. Each speaker made four recordings of the DRT IV test words. One recording by each speaker was then randomly selected and assembled into one of four six-speaker test tapes.

Test Conditions. One randomly selected six-speaker tape was played through each of thirteen modern digital speech communication systems and the output speech recorded. Output recordings were then presented to the listening crew.

Results and Discussion

The analysis of results was addressed to the issue of the consistency of system differences across speakers. The results of this analysis are presented graphically in Fig. 4. In the figure, total DRT scores, averaged for six speakers, are plotted against the total DRT scores for individual speakers.

Two aspects of the plots are of interest. First is the slope of the regression line for each speaker; second is the dispersion of points about each regression line. With regard to the first aspect, it is clear that speakers vary somewhat in terms of absolute sensitivity to the type(s) of degradation involved. Other things equal, speakers BV, JE and SN are somewhat more sensitive to system differences than speakers RD, CH and BL. With regard to consistency of results, however, the situation is somewhat different. Deviations from the indicated regression line tend to be smaller for BV, RD and CH than for the other speakers, which results have important practical implications. Specifically, it would appear that results for these speakers conform most nearly (but for scale factor differences) to the results for the combined speakers. Under circumstances which do not warrant or permit the use of multiple speakers, BV, RD or CH would be the speakers of choice. With appropriate adjustments for scale factors, data obtained from these speakers could be used to predict the average scores that would be obtained for the entire group of speakers. Table 10 is designed to implement this procedure. Presented in the table are equivalent group averages for individual DRT IV scores yielded by each of the six speakers.

It would appear, from the above results, that <u>speaker x system</u> interactive effects, while rather pronounced under extreme laboratory conditions, may be of relatively minor consequence in the practical testing situation -- particularly in the case of tests performed to compare generally similar devices or systems. It is perhaps desirable, however, to use multiple speakers whenever feasible and, moreover, to select the speaker(s) used on the basis of some such criteria as pattern deviation scores obtained under various, representative transmission conditions.

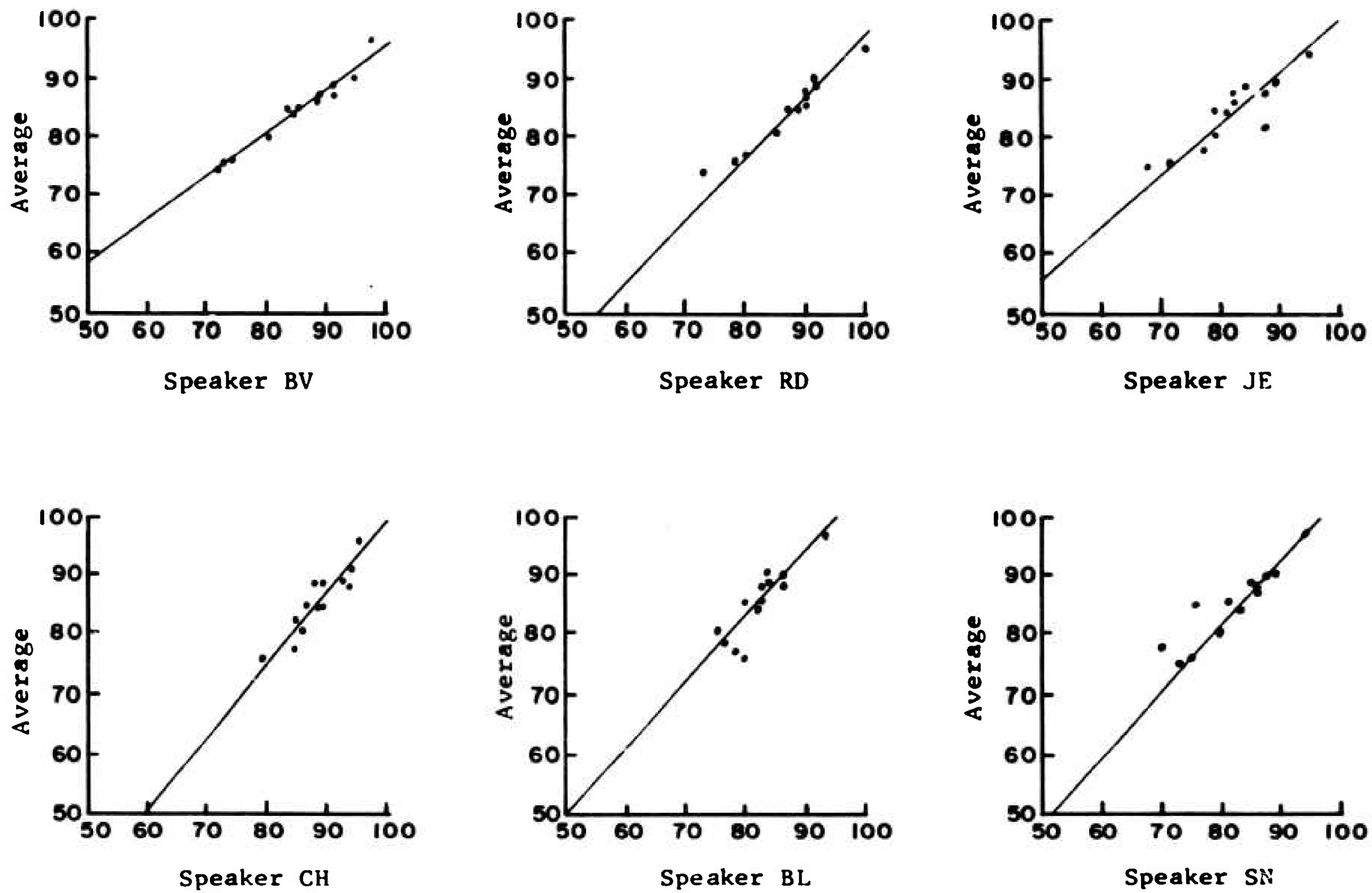

Fig. 4. Averaged DRT Total Scores of Six Speakers Plotted Against DRT Total Scores of Individual Speaker for a Sample of Present-Day Digital Vocoders.

TABLE 10. Equivalencies Between DRT IV Total Scores for Individual Speakers and DRT IV Total Scores as Averaged for Six Speakers.

BV	AVER	RD	AVER	JE	AVER
100.000	97.300	100.000	98.000	100.000	97.000
95.000	93.410	95.000	92.650	95.000	92.500
90.000	89.520	90.000	87.300	90.000	88.000
85.000	85.630	85.000	81.950	85.000	83.500
80.000	81.740	80.000	76.600	80.000	79.000
75.000	77.850	75.000	71.250	75.000	74.500
70.000	73.960	70.000	65.900	70.000	70.000
65.000	70.070	65.000	60.550	65.000	65.500
60.000	66.180	60.000	55.200	60.000	61.000
55.000	62.290	55.000	49.850	55.000	56.500
50.000	58.400	50.000	44.500	50.000	52.000
45.000	54.510	45.000	39.150	45.000	47.500
40.000	50.620	40.000	33.800	40.000	43.000
35.000	46.730	35.000	28.450	35.000	38.500
30.000	42.840	30.000	23.100	30.000	34.000
25.000	38.950	25.000	17.750	25.000	29.500
20.000	35.060	20.000	12.400	20.000	25.000
15.000	31.170	15.000	7,050	15.000	20.500
10.000	27.380	10.000	1.700	10.000	16.000
5.000	23.390	5.000	-3.650	5.000	11.500
.000	19.500	.000	-9.000	.000	7.000

CH	AVER	BL	AVER	SN	AVER
100.000	100.600	100.000	101.600	100.000	102.790
95.000	94.600	95.000	97.015	95.000	97.290
90.000	88.600	90.000	92.430	90.000	91.790
85.000	82.600	85.000	87.845	85.000	86.290
80.000	76.600	80.000	83.260	80.000	80.790
75.000	70.600	75.000	78.675	75.000	75.290
70.000	64.600	70.000	74.090	70.000	69.790
65.000	58.600	65.000	69.505	65.000	64.290
60.000	52.600	60.000	64.920	60.000	58.790
55.000	46.600	55.000	60.335	55.000	53.290
50.000	40.600	50.000	55.750	50.000	47.790
45.000	34.600	45.000	51.165	45.000	42.290
40.000	28.600	40.000	46.580	40.000	36.790
35.000	22.600	35.000	41.995	35.000	31.290
30.000	16.600	30.000	37.410	30.000	25.790
25.000	10.600	25.000	32.825	25.000	20.290
20.000	4.600	20.000	28.240	20.000	14.790
15.000	-1.400	15.000	23.655	15.000	9.290
10.000	-7.400	10.000	19.070	10.000	3.790
5.000	-13.400	5.000	14.485	5.000	-1.710
.000	-19.400	.000	9.900	.000	-7.210

65.

CHAPTER 4

STRUCTURE OF PHONEMIC INFORMATION IN THE ORAL AND NASAL OUTPUTS

by

Alan D. Sharpley

STRUCTURE OF PHONEMIC INFORMATION IN THE ORAL AND NASAL OUTPUTS*

Introduction

In 1968 S. R. Hyde reported a technique that physically isolated the acoustic outputs of the oral and nasal cavities.[1] The technique involved the separation of the two outputs by a metal acoustic shield that was fitted to the speaker's head. Then, while a speaker was fitted into the separation device, the oral and nasal outputs were simultaneously recorded during continuous speech. Hyde's results appeared in the form of the sound spectrograms of the two outputs which he compared to each other as well as to the spectrograms for normal speech.

The present study uses a technique similar to that described by Hyde, but, while his interests lie primarily in the differences among the physical waveforms, the purpose here is to determine the relative contributions of the two outputs to the process of consonant recognition. The Diagnostic Rhyme Test is employed in the present study as a means of evaluating the perceptually significant content of the acoustic outputs of the oral and nasal cavities.

---

*The research described in this report was conducted in partial fulfillment of the requirements for the degree of Master of Arts, the University of Texas, 1970.

Acoustic Shield

An acoustic shield was constructed from four sheets of acoustical fiberboard and two sheets of lead which were notched to fit around the speaker's head. During the actual speech recordings the speaker's head was situated in the center of an 8' x 8' x 1" shield consisting of two layers of fiberboard separated by a layer of lead. In addition to its acoustic insulation properties, the lead provided the structure with enough mass to damp the natural fiberboard resonance. Fit of the shield around the speaker's head was sufficient to prevent significant sound leakage, but not so tight as to alter normal speech articulation. The fiberboard sheets directly under the speaker's nostrils and directly above his mouth were beveled ½ inch in order to minimize obstruction of the breath streams and interference with upper-labial articulation. Finally, the shield structure was suspended from the ceiling of a sound-proofed, anechoic chamber (12' x 12' x 12').

A Bruel and Kjaer free-field microphone and a loudspeaker were connected to a level recorder and a beat frequency oscillator in such a way as to measure the amount of attenuation provided by the shield across the frequency range 50 - 10,000 Hz. Figure 5 shows the attenuation characteristic of the acoustic shield that resulted from this measurement.

Recording Procedures and Materials

Bruel and Kjaer #4131 free-field microphones, #2613 cathode

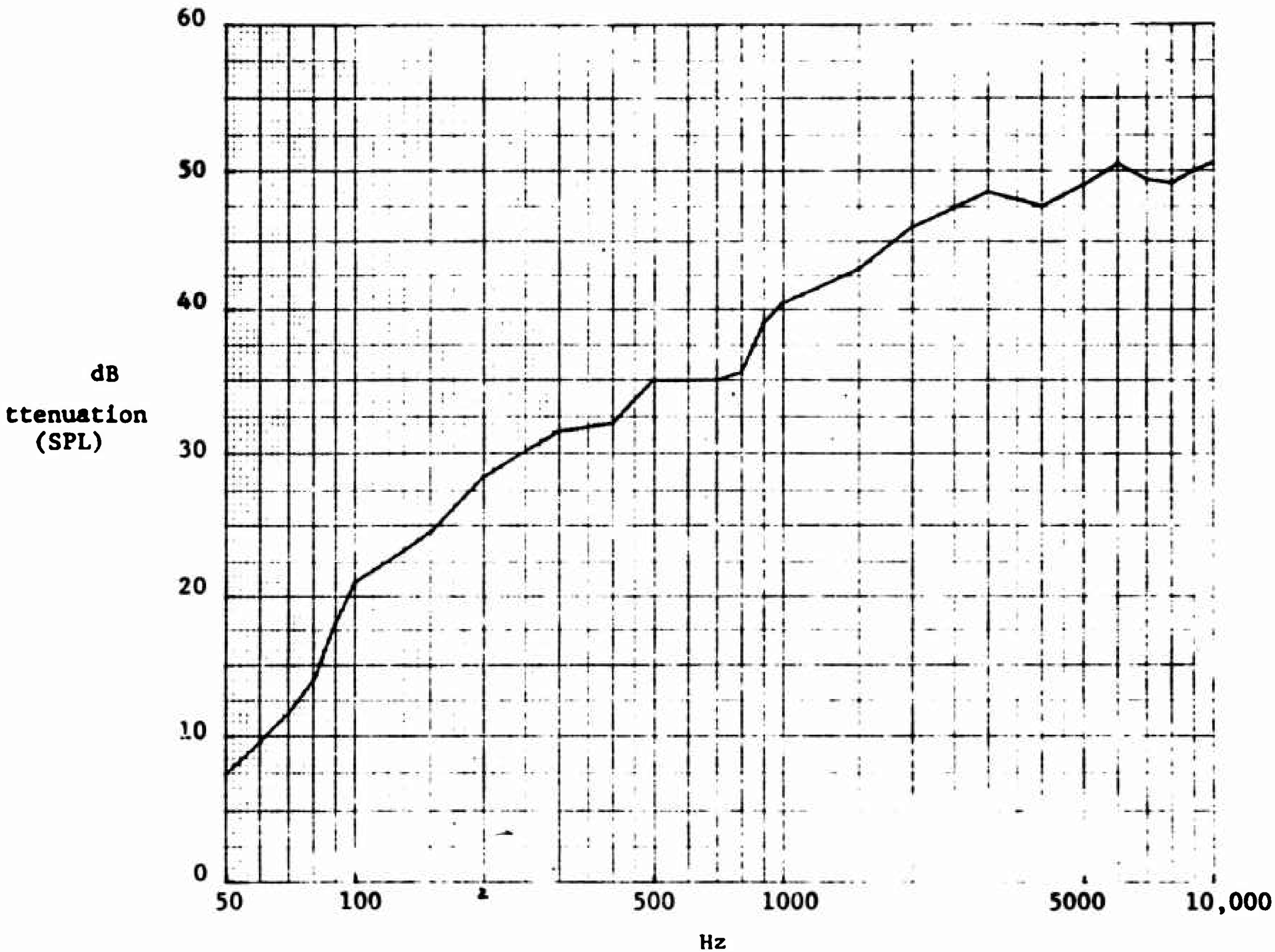

Fig. 5 Attenuation Characteristic of the Acoustic Shield

followers, and #2604 amplifiers were connected to separate channels of an Ampex 602.2 tape recorder. The frequency characteristics of the two microphones were almost identical, being essentially flat in the range 20-10 000 Hz. The separate audio subsystems (microphone, cathode follower, amplifier) were calibrated by pistonphone so that they had virtually identical frequency responses.

The microphones were suspended 10 cms. from their respective sources at 90° incidence, but were kept close to the shield (1 cm.), so that any speech reflecting off the surface of the shield would arrive at the microphones approximately in phase with non-reflected waves. Figure 6 shows a block diagram of the equipment, shield, and chamber used to record the speech material. Figures 7(a) and 7(b) show the configuration of the speaker and acoustic shield in the anechoic chamber during recording of the speech materials in the experimental conditions.

In the control condition, the speaker was situated in the center of the anechoic chamber with his head held firmly by a special restraining device and with a free-field microphone suspended 10 cms. from his mouth at 90° incidence. Speech material was recorded on a single channel of the Ampex tape recorder after passing through one of the microphone/cathode follower/amplifier sub-systems described above.

Two randomizations of the Diagnostic Rhyme Test (DRT)

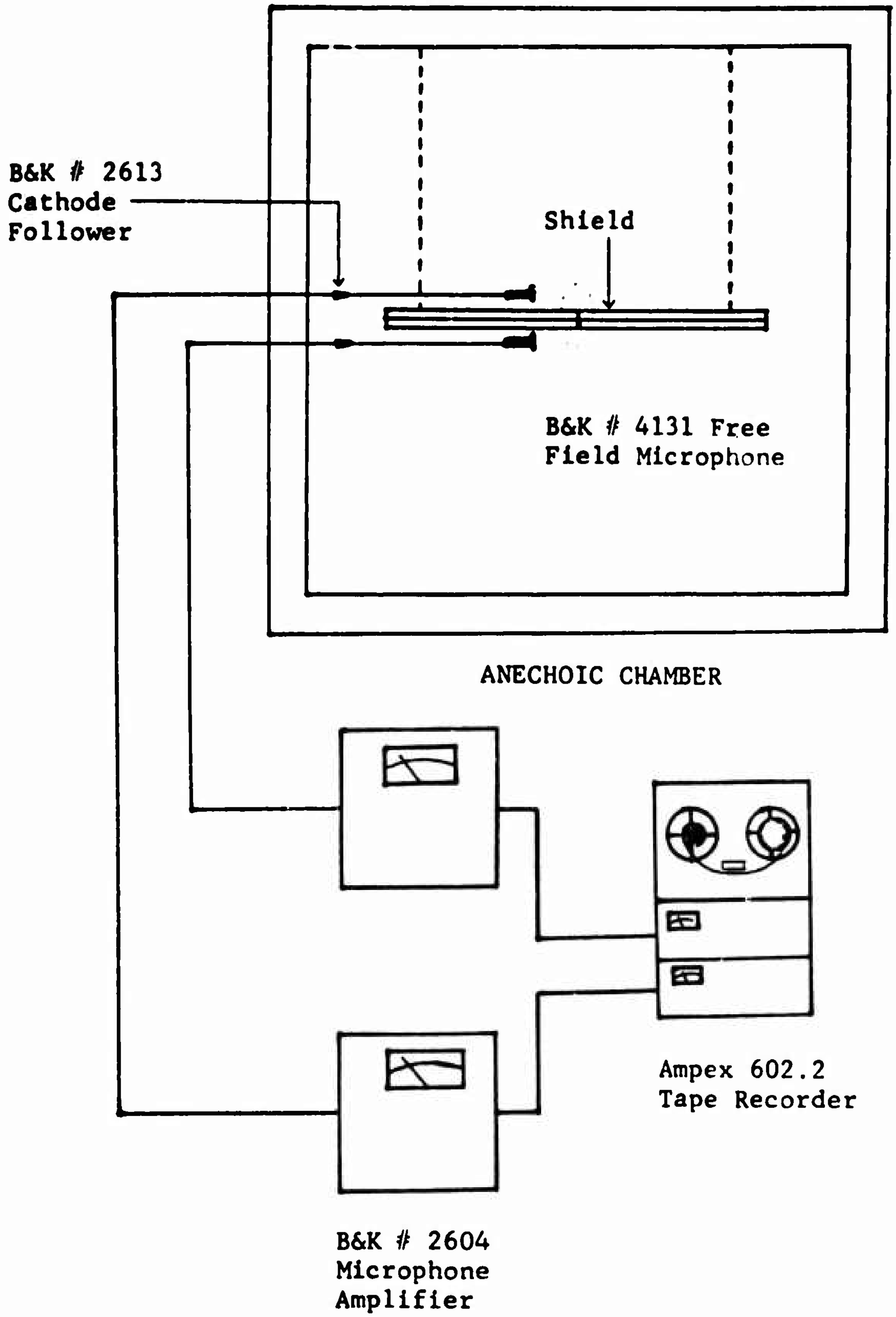

Fig. 6. Diagram of Audio Equipment, Shield, and Chamber Used in Recording of Speech Material.

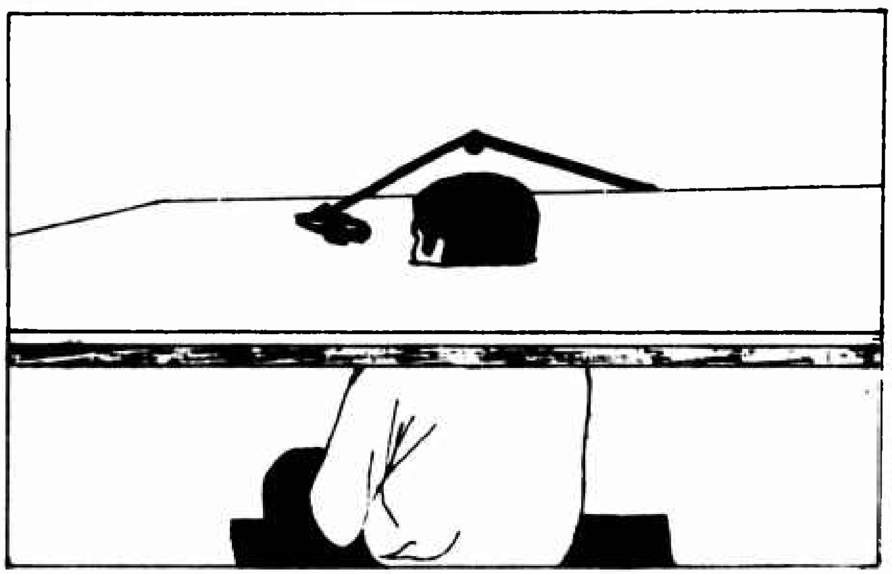

Fig. 7(a). Speaker and Acoustic Shield Situated in the Anechoic Chamber.

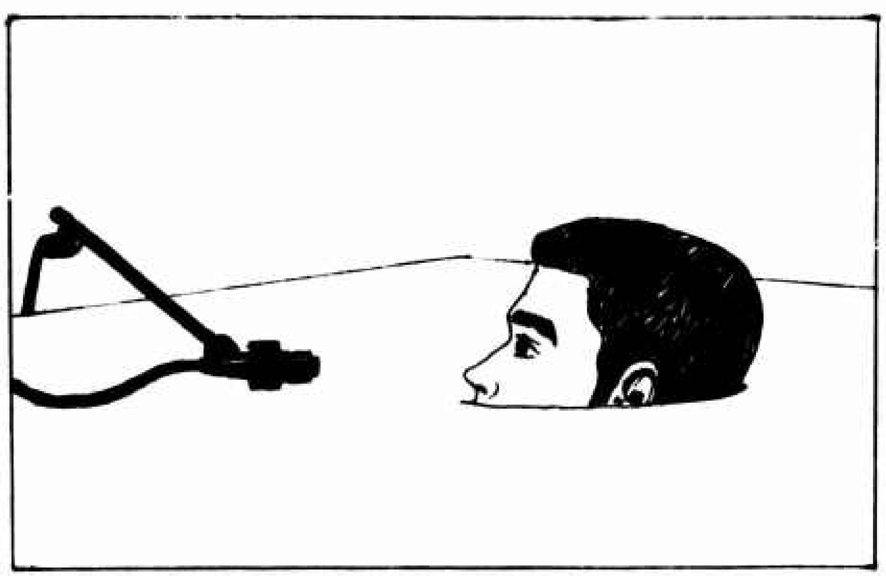

Fig. 7(b). Speaker and Placement of the Nasal Microphone.

materials were recorded under the experimental conditions (two microphones separated by the acoustic shield) and two randomizations under the control conditions (a single microphone). The tapes were edited and 1,000 Hz calibration tones were recorded at the same level as the average of the vowel peaks (VU) for each tape. Since multiple presentations of each randomization were required in the course of the experiment, each basic randomization was partitioned into quarters, which in turn were ordered in various ways to guard against the effects of learning.

Listeners

Eight male University of Texas undergraduates, selected on the basis of consistency of performance on speech intelligibility tests, served as subjects. All had good hearing as determined by pure tone audiometry, and had more than 40 exposures to various intelligibility test materials prior to their participation in the present investigation.

Speaker

A male graduate student at the University of Texas with previous experience as an intelligibility-test speaker provided all recorded speech materials for this investigation. His speech was of General American Dialect and showed no perceptible defects or abnormalities. His intelligibility was highly typical of a large pool of male speakers under a variety of speech transmission conditions.

Presentation Apparatus and Design

Two (6' x 12') double-walled I.A.C. rooms were each partitioned into four listening booths. Speech recordings (DRT) were played on an Ampex 602.2 stereo tape recorder and channeled through a high quality, custom built audio mixer/amplifier, where presentation condition and speech level was determined. The amplified, and, under one condition, mixed, speech was then low-passed at 8,000 Hz by a Krohn-Hite (48 dB/octave) filter, and presented through TDH-39 earphones cased in Rudmose Otocups.

Throughout the experiment, speech level remained constant at approximately 45 dB SPL for the average of the vowel peaks. The experimental design involved four conditions, each representing a different recording mode. Two DRT's were presented under each of the three experimental conditions: nasal output (NO), oral output (OO), electronically mixed nasal and oral outputs (EM), and the single microphone control condition. The eight DRT's were presented in two one-hour testing sessions (four to each) in a counterbalanced arrangement.

Results

Data for the DRT are presented as "percent correct discrimination" scores. For each of the attributes (voicing, nasality, sustention, sibilation, graveness, and compactness) used in the DRT, attribute present, absent, mean, and bias (present - absent) scores, as well as a "Total DRT Score," are presented for each of the experimental conditions.

The "Total DRT Score" may be used as a gross measure of overall intelligibility in that it has been found to correlate highly with scores of other conventional intelligibility tests, i.e., the Fairbanks Rhyme Test.[2] The total DRT Score for the NO condition (24.9) indicates that it was considerably less intelligible than any of the other conditions. It was also found that the total scores for OO (93.8) and EM (94.9) differed non-significantly from the control (95.1) and from each other. Scores for each of six consonant attributes under the three experimental conditions and the control condition are presented in Tables 11 to 14.

The attribute scores for NO are, with a few exceptions, quite low -- many differing non-significantly from chance (Table 11). The exceptions referred to are the mean scores for the "voicing" and the "nasality" attributes. In addition to those scores, the large attribute bias score for "nasality" is notable, i.e., nasal consonants were significantly more distinguishable in NO than were their oral counterparts.

Tables 12, 13, and 14 reveal small differences in the various DRT scores, with the exception of a depressed attribute-present score for "nasality" in the OO experimental condition (Table 2). This decrement in the discriminability of nasal consonants results in a decreased attribute mean score and a relatively large

TABLE 11. Percent Correct Response for the Nasal Acoustic Output (NO).

Consonant Attribute	Attribute Present		Attribute Absent		Attribute Bias*		Attribute Mean	
(Present/Absent)	$\overline{X}$	$s_{\overline{x}}$	$\overline{X}$	$s_{\overline{x}}$	$\overline{X}$	$s_{\overline{x}}$	$\overline{X}$	$s_{\overline{x}}$
Voiced/Voiceless	29.7	10.04	40.6	7.61	-10.9	9.11	35.2	7.65
Nasal/Oral	73.8	10.14	25.4	5.39	48.4	9.38	49.6	6.63
Sustained/Interrupted	19.5	9.95	21.9	5.25	- 2.3	13.15	20.7	4.48
Sibilated/Unsibilated	- 3.5	7.27	27.7	7.02	-31.3	11.63	12.1	4.15
Grave/Acute	12.9	7.62	22.7	8.81	- 9.8	14.22	17.8	4.16
Compact/Diffuse	18.0	6.52	10.5	5.47	7.4	8.43	14.3	4.29

Total DRT Score

$\overline{X}$ = 24.9

$s_{\overline{x}}$ = .73

* Bias score = (Present score)-(Absent score).

TABLE 12. Percent Correct Response for the Oral Acoustic Output (OO).

Consonant Attribute	Attribute Present		Attribute Absent		Attribute Bias		Attribute Mean	
(Present/Absent)	$\bar{X}$	$s_{\bar{x}}$	$\bar{X}$	$s_{\bar{x}}$	$\bar{X}$	$s_{\bar{x}}$	$\bar{X}$	$s_{\bar{x}}$
Voiced/Voiceless	98.0	1.17	95.7	1.95	2.3	1.53	96.9	1.42
Nasal/Oral	83.6	3.58	93.4	1.91	- 9.8	3.09	88.5	2.27
Sustained/Interrupted	92.2	2.29	80.9	3.71	11.3	5.18	86.5	1.67
Sibilated/Unsibilated	99.2	.51	97.7	.98	1.6	.84	98.4	.66
Grave/Acute	92.2	2.89	96.9	1.32	- 4.7	2.57	94.5	1.84
Compact Diffuse	98.4	.84	97.7	1.94	.8	1.94	98.0	1.13

Total DRT Score

$\bar{X}$ = 93.8

$s_{\bar{x}}$ = .97

TABLE 13. Percent Correct Response for the Oral and Nasal Outputs, Electrically Mixed (EM)

Consonant Attribute (Present/Absent)	Attribute Present		Attribute Absent		Attribute Bias		Attribute Mean	
	$\overline{X}$	$s_{\overline{x}}$	$\overline{X}$	$s_{\overline{x}}$	$\overline{X}$	$s_{\overline{x}}$	$\overline{X}$	$s_{\overline{x}}$
Voiced/Voiceless	98.8	.57	98.0	1.01	.8	.78	98.4	.72
Nasal/Oral	96.1	1.84	96.5	1.61	- .4	2.08	96.3	1.38
Sustained/Interrupted	87.1	3.47	76.2	3.01	10.9	5.57	81.6	1.66
Sibilated/Unsibilated	99.2	.51	98.8	.82	.4	.92	99.0	.51
Grave/Acute	94.5	1.42	99.2	.51	- 4.7	1.18	96.9	.89
Compact/Diffuse	97.3	.92	96.5	.92	.8	1.14	96.9	.72

Total DRT Score

$\overline{X} = 94.9$

$s_{\overline{x}} = .73$

TABLE 14. Percent Correct Response for the Output of the Control Condition.

Consonant Attribute	Attribute Present		Attribute Absent		Attribute Bias		Attribute Mean	
(Present/Absent)	$\overline{X}$	$s_{\overline{x}}$	$\overline{X}$	$s_{\overline{x}}$	$\overline{X}$	$s_{\overline{x}}$	$\overline{X}$	$s_{\overline{x}}$
Voiced/Voiceless	97.3	.92	95.3	1.96	2.0	1.86	96.3	1.22
Nasal/Oral	96.5	1.50	98.8	.82	-2.3	1.53	97.7	.93
Sustained/Interrupted	87.9	3.47	82.8	3.29	5.1	6.49	85.4	.93
Sibilated/Unsibilated	99.6	.39	98.4	1.18	1.2	1.17	99.0	.66
Grave/Acute	94.1	2.32	95.7	2.04	-1.6	3.13	94.9	1.53
Compact/Diffuse	98.4	.84	96.1	.78	2.3	1.42	97.3	.39

Total DRT Score

$\overline{X} = 95.1$

$s_{\overline{x}} = .61$

negative attribute bias score for "nasality" in the OO condition.

Differences between the control condition and each of the experimental conditions were evaluated by means of "t"-tests, the results of which are shown in Table 15. This table does not show "t"s for the attribute-present or attribute-absent scores, since that information can be determined from the "t"s for the mean and bias scores (as long as the direction of the bias is known).

Table 15(a) shows the results of "t"-tests between the control and the OO experimental condition. The mean and bias differences for "nasality" are significant ($p < .01$), while all other attributes show "t"s less than 1.0. It appears, then, that the oral cavity produces an output that is significantly deficient in nasality information, but nevertheless retains a substantial amount of information with respect to the state of this feature. A significant difference in bias indicates that the information loss occasioned by removal of the nasal component of the speech signal is an assymetrical loss. Predictably, greatest loss occurs with respect to the positive (i.e., nasal) state of this feature. On the other hand, the output of the nasal cavity presents an entirely different picture.

Table 15(b) reveals the NO condition to be substantially inferior to the control from the standpoint of overall consonant discriminability. The nasal output is deficient in information with respect to all consonant attributes, as seen in the highly

TABLE 15 Results from Analyses by "t"-Tests Between Each of the Three Experimental Conditions and the Control.

		Consonant Attribute (Present/Absent)	"t" for Attribute Mean	"t" for Attribute Bias
(a)	Control-OO	Voiced/Voiceless	.47	.26
		Nasal/Oral	4.89	3.64
		Sustained/Interrupted	.80	.98
		Sibilated/Unsibilated	.55	.24
		Grave/Acute	.27	.76
		Compact/Diffuse	.56	.71
(b)	Control-NO	Voiced/Voiceless	8.22	1.71
		Nasal/Oral	8.36	5.56
		Sustained/Interrupted	14.44	.70
		Sibilated/Unsibilated	21.30	2.72
		Grave/Acute	24.26	.69
		Compact/Diffuse	18.52	.66
(c)	Control-EM	Voiced/Voiceless	1.77	.81
		Nasal/Oral	1.82	1.17
		Sustained/Interrupted	1.99	1.20
		Sibilated/Unsibilated	.00	1.53
		Grave/Acute	1.53	1.02
		Compact/Diffuse	.61	.84

With 7 df, $P < .01$ for "t" ≥ 3.50.

With 7 df, $P < .001$ for "t" ≥ 5.41.

significant ($p < .001$) "t"s for the mean diagnostic scores. Moreover, the significant bias score for "nasality" indicates that a loss, albeit an assymetrical one, occurs even in the case of the feature, nasality. The negative (i.e., non-nasal) state of this attitude is poorly represented in the nasal signal.

Finally, Table 15(c) presents the results of "t"-tests between EM and control. These tests revealed no significant differences between EM and the control in any of the attribute scores. It will be assumed, therefore, that mixing the outputs of the oral and nasal cavities produced a signal that was not significantly different from the control, with respect to consonant discriminability.

DISCUSSION

Since the Total DRT Score represents a measure of speech intelligibility.[*] These scores may be used as an indicator of the relative contribution of the oral and nasal outputs to overall speech intelligibility. The Total DRT Scores for the NO and OO conditions are 24.9% and 93.8% respectively. Predictably, the output of the oral cavity makes a much greater contribution to the speech communication process than does the nasal output.

The OO condition contained sufficient information, relative to the control, to discriminate among consonants with respect to all attributes except "nasality." Even in the case of "nasality," discrimination of the absent state was relatively unimpaired. The oral cavity output was, however, deficient in perceptual information with respect to the state of the feature, "nasality." On the other hand, the NO condition contained little of the information necessary for consonant discrimination on the basis of any of the attributes used in the DRT. In fact, only in the case of "nasality" was there sufficient information for reliable discrimination (49.6%). And even in that attribute the discrimination of orals from nasals was only 25.4% above chance, while the inverse discrimination was 73.8%.

Although there is some information contained in the nasal output for all the consonant attributes, overall speech intelligibility is relatively unimpaired by its absence. In fact, only in the case of the attribute "nasality" and only for the discrimination of that attribute's present state does the nasal output's contribution to consonant discriminability become significant, i.e., the absence of the nasal acoustic output (the 00 experimental condition) results in a significant decrement only in the discrimination of the nasal consonants (/mnŋ/) from their oral cognates (/bdg/).

However, the fact that other attribute mean scores, in addition to that of "nasality," are significantly above chance performance in the NO condition (Table 16) is somewhat remarkable. It seems unlikely, in view of the low level at which the speech was presented to the crew of listeners, that this NO information is simply output from the oral cavity that was not completely attenuated by the acoustic shield. If such were the case, the NO scores would be the result of high frequency distortion, since the acoustic shield served, in effect, as a low-pass filter, as indicated by the graph of Fig. 5. However, the patterns of mean diagnostic scores and bias scores obtained under the NO condition are not characteristic of those which have been found in cases involving low-passed speech,[4] nor do the patterns parallel those that have been obtained for speech presented under low signal-to-noise ratios.[5] It seems, therefore, that the attribute

Table 16. Results from Analysis by "t" Test for the NO Experimental Condition (Significance of Attribute Scores with Respect to Chance Performance).

Consonant Attribute	"t" for Mean Diagnostic Score
Voicing	4.60
Nasality	7.48
Sustention	4.62
Sibilation	2.91
Graveness	4.27
Compactness	3.33
Total DRT Score	5.80

With 7 df, $P < .05$ for "t" ≥ 2.37

With 7 df, $P < .01$ for "t" ≥ 3.50

With 7 df, $P < .001$ for "t" ≥ 5.41

scores that differ significantly from chance in the NO condition are not artifactual, but rather that they result from the actual presence of perceptual discriminatory information in the output of the nasal cavity.

REFERENCES

[1]S. R. Hyde, "Nose Trumpet: Apparatus for Separating the Oral and Nasal Outputs in Speech," Nature, 219, 763-765 (1968).

[2]W. D. Voiers, "Performance Evaluation of Speech Processing Devices, I. Intelligibility, Quality, Speaker Recognizability," Final Report, Contract No. AF19 (628) 4195, AFCRL (1965).

[3]*Ibid*.

[4]W. D. Voiers, paper in preparation.

[5]A. D. Sharpley, "Effects of Presentation Level on Listener Response to the Diagnostic Rhyme Test Materials," in W. D. Voiers, "Diagnostic Evaluation of Speech Processing Systems," Final Report, Contract No. F 19628-68-C-0068, AFCRL, 32-35 (1968).

CHAPTER 5

DIAGNOSTIC EVALUATION OF INTELLIGIBILITY IN PRESENT-DAY DIGITAL VOCODERS

by

William D. Voiers and Caldwell P. Smith

DIAGNOSTIC EVALUATION OF INTELLIGIBILITY IN PRESENT-DAY DIGITAL VOCODERS

William D. Voiers
TRACOR, Inc.
Austin, Texas

Caldwell P. Smith
Data Sciences Laboratory
Laurence G. Hanscom Field
Bedford, Massachusetts

Summary

Recordings of Form IV of the Diagnostic Rhyme Test by six male speakers were used to evaluate the performance of a sample of digital vocoders, all operating in the neighborhood of 2400 bps. The results are compared to those of the 1967 survey and to the case of noise masked speech. Specific strengths and weaknesses of the "typical vocoder" of 1972 are discussed.

Introduction

The purpose of this report is to attempt to characterize the performance of present-day digital vocoders from the standpoint of speech intelligibility. Ideally, it would serve as a sequel to a similar report generated by the survey conducted in conjunction with the 1967 Speech Conference[1], and thus permit an evaluation of the advances in digital vocoder technology that have occurred during the past five years. Regrettably, several factors converged to preclude such an evaluation on any reasonably controlled basis.

The situation is complicated first by the fact that only one of the systems evaluated in the previous survey was available for evaluation in the present survey. In addition, the sample of pitch excited, digital vocoders available for purposes of the present survey was even smaller than the sample used in the previous survey. Among these, moreover, one was clearly malfunctioning to a degree that warranted its exclusion from the survey. Another factor which complicates, but does not in itself preclude, comparisons was the use of a different, albeit improved, version of the Diagnostic Rhyme Test, DRT Form IV[2]. On the positive side are the more refined evaluations permitted by the current version of the DRT, and by the use of multiple speakers, one of whom served as the single speaker used in the earlier survey.

Other things equal, DRT Form IV tends to yield somewhat lower scores than the DRT Form III used in the previous survey. However, results obtained with this form can be rather easily translated into their Form III equivalents. For example, half of the items used in Form IV to test the apprehensibility of the attribute <u>sustention</u> involve voiced consonant pairs, while half involve unvoiced pairs. These proportions are different in the case of Form III, where unvoiced consonant pairs predominate. Since <u>sustention</u> tends generally to be more apprehensible in unvoiced pairs, DRT III typically yields higher scores on the sustention scale than DRT IV. However, by appropriately weighting listener performance on voiced and unvoiced pairs, sustention scores on one form of the DRT can be translated into their equivalents on the other. Similar adjustments can also be made in the case of scores for <u>voicing</u> (where friction is the conditioning factor), <u>graveness</u> (where difficulty is conditional upon the state of <u>voicing</u> and "plosion"), and so on.

Methods and Materials

For purposes of this investigation, six male speakers recorded four complete sets (192 test words each) of the DRT IV materials. These were randomly combined into four master tapes, each of which contained a recording from all six speakers. These were randomly assigned to the various entries in the survey (which included systems other than digital vocoders), but with the restriction that all representatives of a definable class of systems (such as pitch excited digital vocoders) received copies of the same master tape. The output recordings from all entries were presented in random order to a crew of eight highly selected (for stability of performance) and experienced listeners. This order was reversed and all materials were presented a second time to the same crew.

All test materials were presented diotically at an SPL of approximately 72 dB. Proprietary considerations preclude disclosure of the exact number and identities of the systems involved.

Results

It may be of interest, first, to compare the performance of the present sample of vocoders with that of the vocoders evaluated in the previous survey. For this purpose, only data for the single speaker common to the two surveys are used. The averages of the major diagnostic scores yielded by the present sample were translated into their DRT III equivalents. They are presented in Table 17.

Table 17. Equivalent DRT III Scores for Three Conditions

Condition	Vo	Na	Su	Si	Gr	Co	Av
	Diagnostic Scale						
Typical Dig. Vocoder 1967 (DRT III)	97	98	82	97	89	93	93
Typical Dig. Vocoder 1972 (DRT III equiv.)	95	97	83	99	82	94	92

From the table, we can only conclude that the "typical" digital vocoder of 1972 differs negligibly from that of 1967 when evaluated on the basis of essentially the same criteria. The average DRT total score of the present-day sample falls one point below that of the 1967 vocoder. This result, however, merits only the most qualified acceptance, in view of degree of intervocoder variation that characterized both samples. In both 1967 and 1972, total scores spanned a range of over three points. The addition or exclusion of a single case from either sample could easily tilt the balance in favor of one or the other. Finally, some allowance must of course be made for inadequacies in the procedure used for converting DRT IV results to their DRT III equivalents. Although different listener crews were used, this factor would appear to be of negligible consequence. When the present crew was used to evaluate sample tapes from the 1967 survey, differences in total DRT scores were typically of the order of .1 percent.

Table 18 presents the average of the unadjusted diagnostic scores yielded by the present sample of vocoders. For purposes of comparison, corresponding scores for the case of noise masked speech (6 dB S/N ratio, 8 KHz passband) are also presented. The standard errors shown in this table are derived from mean scores for speakers rather than listeners, since the former constitute the more important source of variation in test results.

Table 18. Gross DRT IV Diagnostic Scores for the Typical Vocoder and for Noise Masked Speech

	Condition			
	Vocoded Speech		Noise Masked	
Score	$\overline{x}$*	s.e.**	$\overline{x}$*	s.e.**
Voicing	86	3.2	94	1.0
Nasality	96	1.5	98	0.6
Sustention	73	2.5	76	2.9
Sibilation	96	1.4	95	1.2
Graveness	77	1.9	74	2.4
Compactness	93	0.8	90	0.9
Average	89	1.2	88	0.7

*Averages for six speakers
**Based on speaker averages

From the table, it appears that the effects of vocoding upon speech apprehensibility are grossly quite similar to those of noise, where the two conditions yield approximately the same overall level of speech apprehensibility. In any case, such differences as appear here cannot be safely generalized to the population of male speakers at large.

Table 19 provides a more detailed analysis of the "typical vocoder" of 1972. Shown in the table are averages of the six major diagnostic scores for the vocoders in the present sample. Various components of each of these scores are also shown. For example, the voicing

Table 19. Complete Diagnostic Scores for the Typical Digital Vocoder

Attribute	Pos. State	Neg. State	Bias	$S.E._B$	Average	$S.E._A$
Voicing	83.5	88.8	-5.3	4.93	86.1	3.18
Frictional	72.3	80.0	-7.7	9.60	76.1	6.24
Nonfrictional	94.7	97.6	-2.9	1.94	96.2	1.04
Nasality	94.3	96.9	-2.6	1.64	95.6	1.51
Grave	91.6	95.3	-3.7	2.10	93.5	2.94
Acute	97.1	98.5	-1.4	1.66	97.8	.56
Sustention	73.7	71.6	2.1	2.59	72.7	2.49
Voiced	68.6	61.0	7.6	6.18	64.8	4.01
Unvoiced	78.9	82.2	-3.3	7.66	80.6	3.03
Sibilation	94.5	97.6	-3.1	2.00	96.1	1.38
Voiced	90.8	97.1	-6.4	3.56	93.9	2.08
Unvoiced	98.3	98.1	.2	.68	98.2	.73
Graveness	73.5	79.7	-6.2	6.11	76.6	1.91
Voiced	81.5	88.3	-6.8	5.28	84.9	2.62
Unvoiced	65.6	71.1	-5.5	8.76	68.3	1.58
Plosive	82.7	88.6	-5.9	7.78	85.6	2.30
Nonplosive	64.4	70.8	-6.4	5.80	67.6	3.92
Compactness	95.2	91.5	3.6	1.72	93.4	.76
Voiced	97.6	96.1	1.5	.60	96.8	.50
Unvoiced	92.8	87.0	5.8	3.58	89.9	1.45
Sustained	97.7	92.8	4.8	2.58	95.2	1.29
Interrupted	92.7	90.2	2.5	1.64	91.5	.66
B/M	95.5	94.7	.8	1.16	95.1	.66
B/F	94.9	88.4	6.4	4.00	91.6	1.39

score is broken down into two components representing the apprehensibility, respectively, of the positive and negative states of this attribute. It is broken down additionally into two components representing the gross apprehensibility of voicing in frictional (including affricates) and nonfrictional consonants respectively. Further scores are provided for each state of voicing in each of these two cases. Values in the "bias column" indicate the degree to which listeners favored the positive states of the various attributes. The standard errors for bias and total scores are in all cases based on speaker means and thus provide indications of the susceptibility of the various scores to differences in speaker characteristics.

Although few of the trends suggested by these results are statistically significant, several are worthy of remark, particularly as they coincide or fail to coincide with trends observed under other circumstances. There is, for example, a rather strong indication that voicing is less apprehensible in frictional consonants than in nonfrictional consonants. This trend, which also characterizes unprocessed speech in moderate levels of noise, was evident for all six speakers in the present case. The inflated standard error for the frictional case derives in fact from the extreme degree to which this trend was associated with one of the speakers. The negative bias, which appears here, is not significant, nor is it in the case of noisy, unprocessed speech.

The negative average bias shown in the case of nasality is not significant nor is it consistent with results for other transmission conditions.

On the average, listeners in this investigation were consistently able to apprehend the state of sustention more reliably for unvoiced phonemes than voiced. This trend was observed only for five of the six speakers in this case, but is generally observed in the case of noisy speech.

Sibilation appears to be somewhat less apprehensible in the voiced than in the unvoiced case for the present sample of vocoders. This trend is evident for all six speakers and is also found in the case of noisy speech. Results for five of the six speakers reveal a slight negative bias in the case of sibilation. This bias is generally pronounced in the case of noisy speech.

Although the results in Table 19 suggest a rather consistent negative bias in the case of graveness, this tendency was not associated with all six speakers. No such bias is evident in the case of moderately noisy speech although a pronounced positive bias is found in cases involving higher noise levels. The apprehensibility of graveness clearly varies from voiced to unvoiced phonemes and from plosive to nonplosive. These tendencies are evident under virtually all transmission conditions, and derive in part from the fact that the unvoiced, nonplosive pair, /f-θ/, is involved in four of the most difficult items of the DRT IV.

No significant biases are evident in the case of compactness, but the source state of this attribute proves consistently to be more apprehensible in voiced than in unvoiced phonemes. In vocoded speech, compactness appears to be equally apprehensible in sustained and in interrupted phonemes. However, it is consistently more apprehensible in sustained phonemes in the case of noisy speech.

The back-middle distinction appears slightly less difficult, on the average, than the back-front opposition, in the case of vocoded speech. This trend is not evident with all speakers nor is it found in the case of noise masked speech.

It has often been observed that intelligibility test scores depend significantly on the characteristics of the speaker involved and some degree of speaker dependence was evident in the present case. Table 20 presents average diagnostic scores for each of six speakers used in this investigation. In the table, the speakers are ordered with respect to average pitch frequency. Some correlation between pitch frequency and various DRT scores is evident and, although the present sample is insufficient for purposes of generalization, we have consistently observed this trend with larger samples of speakers. Other things equal, low-pitched speakers yield higher DRT scores than high-pitched speakers on pitch excited vocoder systems. Although this tendency is evident to some degree in several diagnostic dimensions, it is most pronounced in the case of voicing. Here, moreover, there are pronounced speaker differences in characteristic bias. Low-pitched speakers tend to induce a positive bias in the case of voicing while high-pitched speakers are consistently associated with negative biases. Although there were minor differences in the ordering of speaker averages from one system to the next, in no case did a score for a high-pitched speaker exceed that of a low-pitched speaker.

Table 20. Diagnostic Scores for Six Speakers (Average for all Vocoders)

	Diagnostic Score						
Speaker	Vo	Na	Su	Si	Gr	Co	Av
CH(LP)	93	98	76	98	81	93	90
BV(LP)	93	97	72	95	81	96	90
RD(IP)	92	97	80	98	75	94	90
BL(IP)	74	95	69	99	78	94	85
JE(HP)	82	88	63	97	78	93	83
SN(HP)	83	98	76	90	68	90	84

The range of speaker averages for individual systems varied between six and nine percentage points and it is conceivable that some such indicant of system versatility could prove to be of value as a supplementary criterion of system performance. Further research on this issue is needed, however.

Conclusions

In conclusion, the typical digital vocoder of 1972 appears grossly to affect speech apprehension in much the same way as band-limited Gaussian noise. As in 1967, voicing, sustention and graveness constitute the phonemic dimensions in which the greatest opportunities for improvement exist. It is evident that the present-day vocoder does not do all things equally well when operating in the voiced and unvoiced modes. In general, it would seem to preserve information as to type of articulation most effectively in the unvoiced state; place of articulation most effectively in the voiced state. It is also evident that present-day vocoders do not perform equally well for all speakers. Low-pitched speakers tend to yield higher scores than high-pitched speakers and other speaker factors will undoubtedly emerge from the results of further research on this issue.

References

[1] W. D. Voiers, "The Present State of Digital Vocoding Technique: A Diagnostic Evaluation", IEEE Transactions on Audio and Electroacoustics, Vol. AU-16, 2, 275-279, 1968.

[2] W. D. Voiers, "Diagnostic Approach to the Evaluation of Speech Intelligibility", presented at the meeting of Groupement des Acousticiens de Langue Francaise, Groupe "Communication Parlee", April, 1971.

CHAPTER 6

SUMMARY OF ACTIVITIES

SUMMARY OF ACTIVITIES

Summarized here are the major accomplishments of the Psychometrics Department, Environment and Physical Sciences Division of TRACOR, Inc., under Contract No. F 19628-70-C-0182.

Publications

Voiers, William D. and Smith, Caldwell P., Diagnostic Evaluation of Intelligibility of Present-Day Digital Vocoders, AFCRL-72-0120, 22 February, 1972, Special Reports, No. 131, 170-175 (IEEE Cat. No. 72) CHO 596-7 AE.

Presentations

Voiers, William D., Current Status of the Diagnostic Rhyme Test, 81st Meeting, Acoustical Society of America, Washington, D. C., April 1971.

Sharpley, Alan D., Structure of Phonemic Information in the Nasal Output, 81st Meeting, Acoustical Society of America, Washington, D. C., April 1971.

Voiers, William D., and Smith, Caldwell P., Diagnostic Evaluation of Intelligibility of Present-Day Digital Vocoders, 1972 Conference on Speech Communication and Processing, Cambridge, Mass., April 1972.

Meetings Attended

80th Meeting of the Acoustical Society of America, Houston, Texas, November, 1970 -- W. D. Voiers, A. D. Sharpley, C. J. Hehmsoth.

81st Meeting of the Acoustical Society of America, Washington, D. C., April, 1971 -- W. D. Voiers and A. D. Sharpley.

1972 Conference on Speech Communication and Processing, Cambridge, Mass., April, 1972 -- W. D. Voiers.

Technical Personnel

Dr. William D. Voiers, Director, Psychometrics Department: Program Manager and Principal Investigator.

Mr. Alan D. Sharpley, Engineer Scientist, Psychometrics Department: Project Engineer.

Mr. Carl J. Hehmsoth, Engineer Scientist, Psychometrics Department.

Research Activities

Approximately half of the effort devoted to this project was directed to the end of developing and validating improved methods of evaluating speech communication systems from the standpoint of intelligibility. This effort culminated with the development and validation of the Diagnostic Rhyme Test Form IV (DRT-IV). It also occasioned research on a diversity of subjects in the area of speech perception, and several of the projects undertaken yielded results of general practical and theoretical interest.

In addition to the design and validation of DRT IV itself were studies of individual differences in speech perception, studies of speaker differences, a study of the information content of the nasal output, and a comparative evaluation of present-day speech communication and processing devices, as reported in Chapters 1-5.

Testing Services

Pursuant to the provisions of the contract, a series of Diagnostic Rhyme Tests were performed on tapes of experimentally processed speech materials supplied by the contract monitor. These included among others output tapes from the various speech communication and processing systems submitted for evaluation in conjunction with the 1972 Conference on Speech Communication and Processing held at Newton, Mass., under the joint sponsorship of the Air Force Cambridge Research Laboratories and the Institute of Electrical and Electronics Engineers, Inc.

Software Development

Analysis data yielded by the investigative phases of the program necessitated the development of a series of successively refined computer scoring programs for use with the Diagnostic Rhyme Test. Such programs make feasible a variety of scoring refinements in the routine use of the DRT for purposes of system evaluation. Programs developed for use in this project were also modified to permit their use with computer systems other than those available at TRACOR. Appendix II contains the basic DRT IV scoring program and a sample printout.

Tape Recordings

The investigations performed under the contract necessitated the assembly of an extensive library of recorded speech materials. This library included recordings of DRT III-A materials and samples of continuous speech for 80 male speakers. These served, among other things, the purposes of research which led to the development of Form IV of the DRT. Recordings of DRT IV materials were also made by a number of speakers. These were used for purposes of research and testing during the later stages of the project. They also provided the basic test materials used in the survey of speech processing and communication systems conducted in conjunction with the 1972 Conference on Speech Communication and Processing. All master recordings were delivered to the contract monitor.

APPENDIX I

FORTRAN MAIN ROUTINE AND SUBROUTINE LISTINGS FOR DRT IV COMPUTER SCORING PROGRAM

FORTRAN Main Routine and Subroutine Listings for DRT IV Computer Scoring Program

MAIN ROUTINE


```
      INTEGER CODE
      COMMON/ALL/CODE(150),SEL(20),NQ,NL,NA,NV,JTTEST,NDL,NASCOR,JPUNCH
     1,NR,JAVE,IBIG,JSPK,IOP(10),ISAVE(200),NATE(20)
      COMMON /MF/ LABLE(13),ITST
      COMMON/MCE/KEY(200,4),JIA,KSAVE(150),NAME(100,2),IANXX(20,4),ISPK
     1EY(10)
      COMMON/ITM/KSPL(11,2,2,2,2),INRD(112,3),LOBLE(9,6),ITOM(112,4),IXS
     1(112,20),ITIM(112,4),IXSB(112,20)
      COMMON/FEW/NNL,NNQ,IKEY(10),LBL(10),LKEY(12,12,12),LCNF(15,6)
      COMMON /FN/ ITEXT (130,6)
      NLKEY=12
      DO 703 J=1,6
703   READ 9001,(ITEXT(L,J),L=1,78)
9001  FORMAT (13A6)
      READ 120,((INRD(I,J),J=1,3),I=1,112)
120   FORMAT (4(3A6,2X))
      READ 121,((LOBLE(I,J),J=1,6),I=1,9)
121   FORMAT(6A6,4X,6A6)
      DO 122 I=1,11
122   READ 123,((((KSPL(I,J,K,L,M),J=1,2),K=1,2),M=1,2),L=1,2)
123   FORMAT (2X,8I4,4X,8I4)
      READ 125,(IKEY(J),J=1,5)
125   FORMAT(6(O12,1X))
      DO 126 I=2,NLKEY
      II=I-1
126   READ 127,((LKEY(I,J,K),K=1,I),J=1,II)
127   FORMAT(40I2)
      READ 121,(LBL(I),I=1,6)
      READ 850,((LCNF(I,J),J=1,6),I=1,15)
850   FORMAT(6A6,4X,6A6)
      READ 109, NKEY,NA,NV,NASCOR,JSA6,JSA7,NDL
      J2PAGE=0
      DO 101 I=1,NKEY
101   READ 110, N,(KEY(N,J),J=1,4)
      READ 113,(ISPKEY(I),I=1,9)
      DO 102 I=5,20
102   READ 108, (IANXX(I,J),J=1,4)
      DO 103 I=1,20
103   READ 111, N,(NAME(N,J),J=1,2)
104   READ 109,NTEST,NQ,NL,JPUNCH,JAVE,NXPAGE,LBLTST,IBIG,JIA,NAKEY,
     1NXLIS,JSPK,JTTEST,(IOP(I),I=1,10)
      IF(JAVE.EQ.0) JAVE=1
      IXZP=0
      IF(NTEST.LT.0) IXZP=2
      IF(MOD(JAVE,JSPK).NE.0) GO TO 300
      GO TO 301
300   PRINT 9052
9052  FORMAT (10X,'****** THE NUMBER OF SPEAKERS AND TESTS ARE INCOMPATIB
     1LE ******')
      GO TO 107
301   CONTINUE
      MLL=NL
      MLQ=NQ
      MLAV=JAVE
      MLSP=JSPK
      IF (NQ.EQ.0) GO TO 107
      PRINT 9049
9049  FORMAT('1  #TS   #QS   #LS   RAT?  #AVE  PAGE  LBL?  BGIO  ITM?  X
```

```
      1KY?  XLS?  SPK?  TTST?   SPLT OPTNS*)
       PRINT 9050,NTEST,NQ,NL,JPUNCH,JAVE,NXPAGE,LBLTST,IBIG,JIA,NXKEY,
      1NXLIS,JSPK,JTTEST,(IOP(I),I=1,10)
 9050  FORMAT (13I6,4X,10I1)
       J2PAGE=0
       IF (NXPAGE) 114,116,116
 114   J2PAGE=1
       NXPAGE=IABS(NXPAGE)
 116   CONTINUE
       J3PAGE=0
       IF(NXPAGE.LT.100) GO TO 117
       J2PAGE=1
       J3PAGE=1
       NXPAGE=NXPAGE-100
 117   CONTINUE
       NTEST=IABS(NTEST)
       NR=0
       IF (JPUNCH.GT.0) NR=6
       IXRA=ISUB(2HAA)
       IF(NXKEY.EQ.0) GO TO 222
 221   READ 9060,IIS,ITOBE,KOP
       ISAVE(IIS)=ITOBE
       IF(KOP.EQ.0) GO TO 221
 9060  FORMAT (I4,2X,I4,65X,I5)
 222   CONTINUE
       IF (NXLIS.NE.0) IXRA=ISUB(2HEX)
       IF(JSPK.EQ.0) GO TO 234
       JSPK=IABS(JSPK)
       DO 231 KZ=1,JSPK
 231   READ 232,L,NATE(L),(NAME(L,LL),LL=1,2)
 232   FORMAT(I4,2X,A2,2X,2R6)
 234   CONTINUE
       IF (JAVE.EQ.0) JAVE=1
       DO 105 J=1,13
 105   LABLE(J)=6H
       IF (LBLTST.EQ.1) READ 112, (LABLE(J),J=1,13)
       DO 106 ITST=1,NTEST
       NL=MLL
       NQ=MLQ
       JAVE=MLAV
       JSPK=MLSP
       LBL(3)=6HFRCT *
       LBL(4)=6HGRAV *
       CALL ITTIAL
       NNQ=NQ
       NNL=NL
       IF(JSPK.NE.0)  NNL=JSPK
       IF (LBLTST.EQ.2) READ 112, (LABLE(J),J=1,13)
       IXRA=ISUB(2HBX)
       CALL RATING (1)
       CALL CHECK
       DO 106 LXZP=1,IXZP
       CALL STOPC
       DO 115 NPAG=1,NXPAGE
       IF(J2PAGE.EQ.1) CALL FINII
       IF(J3PAGE.EQ.0) CALL FINIS
 115   CONTINUE
       CALL ERROR
       IF(IBIG.GT.0) CALL BIGIO
       IF(JIA.GT.0) CALL ITMANL
       IF(JIA.LT.2) GO TO 119
       DO 118 I=1,112
       DO 1018 J=1,4
 1018  ITOM(I,J)=ITIM(I,J)
       DO 1019 J=1,20
 1019  IXS(I,J)=IXSB(I,J)
 118   CONTINUE
       CALL ITMANL
 119   CONTINUE
       IF(JPUNCH.EQ.2) CALL PUNCH
       IF (JTTEST.GT.0) CALL TTST
 106   CONTINUE
       GO TO 104
 107   CONTINUE
 C
 108   FORMAT (10X,O10,5X,O10,5X,O10,5X,O10)
 109   FORMAT (13I4,16X,10I1)
 110   FORMAT (2X,I2,6X,O10,5X,O10,5X,O10,5X,O10)
 111   FORMAT (I4,2X,3A6)
 112   FORMAT (13A6)
 113   FORMAT(10(O6,1X))
       END
```

```
      SUBROUTINE BIGIO
      COMMON/ALL/CODE(150),SEL(20),NQ,NL,NA,NV,JTTEST,NDL,NASCOR,JPUNCH
     1,NH,JAVE,IBUG,JSPK,IOP(10),ISAVE(200),NATE(20)
      COMMON/ERR/NSUB(20),IL(20,7),IX(20,7),IB(20,7),IT(20,7)
     1,IUU(20,150),ITEM(112)
      COMMON/ITM/KSPL(11,2,2,2,2),INRD(112,3),LOBLE(9,6),ITOM(112,4),IXS
     1(112,20)
      DIMENSION IBIG(15),IBIGP(15),BIG(15)
      DO 2 K=1,15
      IBIG(K)=ITEM(1)
      IBIGP(K)=1
      DO 1 I=2,112
      IF(IBIG(K).GT.ITEM(I)) GO TO 1
      IBIG(K)=ITEM(I)
      IBIGP(K)=I
    1 CONTINUE
      J=IBIGP(K)
      ITEM(J)=0
    2 CONTINUE
      TOT=NQ*NL*JAVE
      IF(JSPK.NE.0) TOT=NQ*JAVE*JSPK
      DO 3 K=1,15
      I=IBIGP(K)
      ITEM(I)=IBIG(K)
    3 BIG(K)=((TOT-2*IBIG(K))/TOT)*100.0
      PRINT 10
   10 FORMAT(1H1,9X,56HFOR THE PURPOSES OF FURTHER RESEARCH DESIGNED TO
     1IMPROVE,/,12X,60HYOUR SYSTEM OR DEVICE, YOU WILL FIND IT ADVANTAGE
     2OUS TO GIVE,/,12X,73HSPECIAL ATTENTION TO THE DISTINGUISHABILITY O
     4F THE FOLLOWING WORD PAIRS.           )
      PRINT 16
   16 FORMAT (16X,10HWORD PAIRS  ,17X,4HP(C)  /)
      DO 11 I=1,10
      J=IBIGP(I)
      PRINT 12,(INRD(J,K),K=1,3),BIG(I)
      IF(BIG(I).GT.99.999999) GO TO 20
   11 CONTINUE
   20 CONTINUE
   12 FORMAT (16X,3A6,4X,F10.1,/)
      PRINT 15
   15 FORMAT (/,10X,'**   THE CONTRASTS: FAD-THAD, FIN-THIN,FOUGHT-THOUG
     1HT,',/,12X,'VON-BON, VOX-BOX, VEE-BEE, VILL-BILL, VAULT-FAULT '
     2,/,       12X,                   55HARE GENERALLY AMOUNG THE MOST DI
     5FFICULT TO DISTINGUISH./,12X,74HTHEIR PRESENCE ON THE FOREGOING LI
     7ST DOES NOT, THEREFORE, REFLECT UNIQUELY,/,12X,46HUPON THE PERFORM
     9ANCE OF YOUR SYSTEM OR DEVICE.          )
      RETURN
      END
```

```
      FUNCTION IPOP (I,J)
      JX=J
      JX=JX+28
    1 JX=JX-28
      IF (JX.GT.28) GO TO 1
      JX=JX-1
      IX=I
      IX=LSHIFT(IX,-JX)
      IPOP=AND(IX,1)
      RETURN
      END
```

```
      SUBROUTINE FOOTNT (N)
      COMMON /FN/ ITEXT (130,6)
      GO TO (1,9,1,1,9,1,9,9,1),N
    1 IF(N.EQ.9) J=6
      IF(N.LT.7) J=5
      IF(N.LT.5) J=N
      PRINT 9000,(ITEXT(L,J),L=1,78)
 9000 FORMAT (20X,13A6)
    9 RETURN
      END
```

```
      SUBROUTINE CHECK
      INTEGER CODE
      COMMON/ALL/CODE(150),SEL(20),NQ,NL,NA,NV,JTTEST,NDL,NASCOR,JPUNCH
     1,NR,JAVE,IBIG,JSPK,IDP(10),ISAVE(200),NATE(20)
      COMMON /ERR/ NSUB(20),ILXAP(20,7),ILXAA(20,7),ILXAB(20,7),ILXAT(20
     1,7),ILXQ(20,150),ITEM(112),ISPLT(10,20,2,4),ILXV(20,8)
     2,IWRS(150)
      COMMON/MCE/KEY(200,4),JIA,KSAVE(150),NAME(100,2),IANXX(20,4),ISPK
     1EY(10)
      COMMON /MF/ LABLE(13),ITST
      COMMON/ITM/KSPL(11,2,2,2,2),INRD(112,3),LOBLE(9,6),ITOM(112,4),IXS
     1(112,20),ITIM(112,4),IXSB(112,20)
      DIMENSION IDATA(112)
      JX=0
      DO 4 LL=1,NL
      DO 1 K=1,9
      DO 1 NP=1,2
      DO 1 NZ=1,2
    1 ISPLT(K,LL,NP,NZ)=0
      DO 2 K=1,NA
      ILXAP(LL,K)=0
    2 ILXAA(LL,K)=0
      DO 30 K=1,8
   30 ILXV(LL,K)=0
      DO 3 K=1,150
      IWRS(K)=0
      CODE(K)=4H
    3 ILXQ(LL,K)=0
    4 CONTINUE
      DO 5 I=1,4
      DO 5 J=1,4
    5 IANXX(I,J)=0
      DO 6 J=1,112
      ITEM(J)=0
    6 CONTINUE
      INC=1
      IV=-1
      GO TO 8
    7 INC=-1
      IV=-2
    8 CONTINUE
      DO 23 KY=1,JAVE
      IF (JPUNCH.NE.0) CALL RATING (IV)
      DO 23 KQ=1,NQ
      DO 22 LL=1,NL
      DO 22 MJL=1,2
      READ 26, KODE,IL,IPAGE,NKEY,(IDATA(J),J=1,56),IANS,ISHEET
      IANS=IANS+1
      ISHEET=ISHEET-IRA+1
      IANS=((IANS*4)-4)+((ISHEET+1)/2)
      IF(ISAVE(NKEY).NE.0) NKEY=ISAVE(NKEY)
      IQ=MOD(NKEY,NQ)
      IF(IQ.EQ.0) IQ=NQ
      CODE(KY)=KODE
      IF(JSPK.EQ.0) IQ=(KY*NQ-NQ)+IQ
      IF(JSPK) 76,77,76
   76 IL=-IL
   77 ILX=ISUB(IL)
      NSUB(ILX)=IL
      IL=ILX
      JJ=0
      IF (MOD(ISHEET,2).EQ.0) JJ=56
      DO 22 N=1,8
      NH=N
      IF (MOD(ISHEET,2).EQ.0) NH=N+8
      DO 22 K=1,NA
      J=((N*NA)-NA)+K
```

```
      IITEM=J+JJ
      KLINK=(IITEM+27.0)/28.0
      IF (IPOP(IANXX(IANS,KLINK),J).EQ.1) IDATA(J)=3-IDATA(J)
      GO TO (10,9),IPAGE
 9    IDATA(J)=3-IDATA(J)
 10   IF (IDATA(J)-(IPOP(KEY(NKEY,KLINK),J)+1)) 19,22,11
 11   ILXAP(IL,K)=ILXAP(IL,K)+INC
      LQ=MOD(IQ,NQ)
      IF(LQ.EQ.0) LQ=NQ
      ITIM(J+JJ,LQ)=ITIM(J+JJ,LQ)+INC
      IXSB(J+JJ,IL)=IXSB(J+JJ,IL)+INC
      IJ=1
 12   GO TO (13,13,13,13,14,15,17),K
 13   L=IPOP(ISPKEY(K),NH)+1
      ISPLT(K,IL,L,IJ)=ISPLT(K,IL,L,IJ)+INC
      GO TO 18
 14   L=IPOP(ISPKEY(K),NH)+1
      ISPLT(K,IL,L,IJ)=ISPLT(K,IL,L,IJ)+INC
      KK=K+1
      L=IPOP(ISPKEY(KK),NH)+1
      ISPLT(KK,IL,L,IJ)=ISPLT(KK,IL,L,IJ)+INC
      GO TO 18
 15   KK=6
 16   KK=KK+1
      L=IPOP(ISPKEY(KK),NH)+1
      ISPLT(KK,IL,L,IJ)=ISPLT(KK,IL,L,IJ)+INC
      IF (KK.LT.9) GO TO 16
      GO TO 18
 17   CONTINUE
 18   IF (NASCOR-K) 21,20,20
 19   ILXAA(IL,K)=ILXAA(IL,K)+INC
      LQ=MOD(IQ,NQ)
      IF(LQ.EQ.0) LQ=NQ
      ITIM(J+JJ,LQ)=ITIM(J+JJ,LQ)-INC
      IXSB(J+JJ,IL)=IXSB(J+JJ,IL)-INC
      IJ=2
      GO TO 12
 20   ILXQ(IL,IQ)=ILXQ(IL,IQ)+INC
      JTX=MOD(J,8)
      IF(JTX.EQ.0) JTX=8
      ILXV(IL,JTX)=ILXV(IL,JTX)+INC
      IF(JIA.EQ.0) GO TO 21
      INK=INC
      LQ=MOD(IQ,NQ)
      IF(LQ.EQ.0) LQ=NQ
      ITOM(J+JJ,LQ)=ITOM(J+JJ,LQ)+INK
      IXS(J+JJ,IL)=IXS(J+JJ,IL)+INK
 21   ITEM(J+JJ)=ITEM(J+JJ)+INC
      IGRS(KY)=IGRS(KY)+INC
 22   CONTINUE
 23   CONTINUE
      IF ((JTTEST.GT.0).AND.(INC.EQ.1)) GO TO 7
      DO 25 I=1,NL
      DO 24 J=1,NA
      ILXAT(I,J)=ILXAP(I,J)+ILXAA(I,J)
 24   ILXAB(I,J)=ILXAP(I,J)-ILXAA(I,J)
      DO 25 J=1,9
      DO 25 NP=1,2
      ISPLT(J,I,NP,3)=ISPLT(J,I,NP,1)-ISPLT(J,I,NP,2)
      ISPLT(J,I,NP,4)=ISPLT(J,I,NP,1)+ISPLT(J,I,NP,2)
 25   CONTINUE
      RETURN
 26   FORMAT (R4,I2,I1,I2,2X,28I1,1X,28I1,10X,I1,R1)
      END
```

```
      SUBROUTINE ITTIAL
      COMMON/ITM/KSPL(11,2,2,2,2),INRD(112,3),LORLE(9,6),ITOM(112,4),IXS
     1(112,20)
      DO 10 I=1,112
      DO 5 J=1,20
 5    IXS(I,J)=0
      DO6 J=1,4
 6    ITOM(I,J)=0
 10   CONTINUE
      RETURN
      END
```

```
      SUBROUTINE ERROR
      INTEGER CODE
      COMMON/ALL/CODE(150),SEL(20),NXQ,NL,NA,NV,JTTEST,NDL,NASCOR,JPUNCH
     1,NR,JAVE,IBIG,JSPK,IOP(10),ISAVE(200),NATE(20)
      COMMON /ERR/ NSUB(20),ILXAP(20,7),ILXAA(20,7),ILXAB(20,7),ILXAT(20
     1,7),ILXQ(20,150),ITEM(112),ISPLT(10,20,2,4),ILXV(20,8)
     2,IQRS(150)
      COMMON/MCE/KEY(200,4),JIA,KSAVE(150),NAME(100,2),IANXX(20,4),ISPK
     1EY(10)
      DIMENSION OE(20,150),IT1(150),IT2(150),CHI(20),PCHI(20)
      NQ=NXQ
      K2=6H TEST
      K3=6H
      IF(JAVE.GT.1) K3=6HS
      K1=6H WHOLE
      IF(NQ.EQ.2) K1=6H HALF
      MWOZ=6HLISTEN
      NWOZ=6HERS ON
      IF(JSPK) 48,49,48
48    MWOZ=6HSPEAKE
      NWOZ=6HRS  ON
49    PRINT 41,NL,MWOZ,NWOZ,JAVE,K1,K2,K3
      TOT=0.
      RXX=NASCOR
      QXX=NQ*JAVE
      IF(JSPK.NE.0) QXX=NQ*JSPK*(JAVE/NL)
      TXX=16.*QXX*RXX
      TOT=0.0
      DO 6 I=1,112
      IF (MOD(I,7).EQ.0) GO TO 6
      TOT=TOT+ITEM(I)
6     CONTINUE
      Y=NASCOR*16.
      IF(JSPK.NE.0)  Y=Y*JSPK
      NQ=NXQ*JAVE
      IF(JSPK.NE.0) NQ=NXQ
      DF=NQ-1
      DO 7 I=1,NQ
7     IT2(I)=0
      DO 8 J=1,NL
      IB=KSUM(ILXQ,NQ,-J,NDL)
      CHI(J)=0.0
      I1=KSUM(ILXQ,NQ,-J,NDL)
      E=(Y*I1)/(NQ*Y)
      DO 8 I=1,NQ
      IT2(I)=IT2(I)+ILXQ(J,I)
      IC=KSUM(ILXQ,NL,I,NDL)
      OE(J,I)=ILXQ(J,I)-(IC*IB/TOT)
      CHI(J)=CHI(J)+(((ILXQ(J,I)-E)**2)/E)+(((ILXQ(J,I)-E)**2)/(Y-E))
8     PCHI(J)=SGNF(CHI(J),-DF)
      PRINT 25,(CODE(I),IQRS(I),I=1,JAVE)
      DO 13 I=1,NL
      QXX=KSUM(ILXQ,NQ,-I,NDL)
      RXX=(TXX-2*QXX)/TXX
      IF (JTTEST.GT.0) RXX=QXX/TXX
      RXX=RXX*100
      K=NSUB(I)
      IF(NQ.LE.2) GO TO 37
      PRINT 39,NSUB(I),(NAME(K,L),L=1,2),RXX,SEL(I),CHI(I),PCHI(I),
     1(ILXQ(I,J),J=1,4 ),(OE(I,J),J=1,4 )
39    FORMAT (I4,2A6,3F5.1,R5,2X,4I4,2X, 4F6.1)
      GO TO 3333
37    PRINT 38,NSUB(I),(NAME(K,L),L=1,2),RXX,SEL(I),CHI(I),PCHI(I),
     1(ILXQ(I,J),J=1,2),(OE(I,J),J=1,2)
38    FORMAT (I4,2A6,3F5.1,R5,2X,2I4,10X,2F6.1)
3333  CONTINUE
      IF(NQ.LE.4) GO TO 13
      DO 813 L=5,NQ,4
```

```
      K=L+3
      IF(K.GT.NQ) GO TO 3334
  813 PRINT 40,(ILXQ(I,J),J=L,K),(OE(I,J),J=L,K)
   40 FORMAT (38X,4I4,2X,4F6.1)
      GO TO 13
 3334 K=L+1
      PRINT 3337,(ILXQ(I,J),J=L,K),(OE(I,J),J=L,K)
 3337 FORMAT (38X,2I4,10X,2F6.1)
   13 CONTINUE
      DO 14 J=1,NQ
      IT1(J)=0
      DO 14 I=1,NL
      IT1(J)=ILXQ(I,J)+IT1(J)
   14 CONTINUE
      PRINT 3338,(IT1(J),J=1,NQ)
 3338 FORMAT (38X,4I4)
      NQ=NXQ
      PRINT 29
      DO 16 I=1,NL
      IT1(I)=KSUM(ILXAP,NASCOR,-I,NDL)
      IT2(I)=KSUM(ILXAA,NASCOR,-I,NDL)
   16 PRINT 30, NSUB(I),(ILXAP(I,J),J=1,NA),IT1(I),(ILXAA(I,J),J=1,NA),I
     1T2(I)
      DO 17 J=1,NA
      IT1(J)=KSUM(ILXAP,NL,J,NDL)
   17 IT2(J)=KSUM(ILXAA,NL,J,NDL)
      I1=KSUM(IT1,NASCOR,1,1)
      I2=KSUM(IT2,NASCOR,1,1)
      PRINT 31, (IT1(J),J=1,NA),I1,(IT2(J),J=1,NA),I2
      PRINT 27
      DO 18 I=1,NL
      IT1(I)=KSUM(ILXAB,NASCOR,-I,NDL)
      IT2(I)=KSUM(ILXAT,NASCOR,-I,NDL)
   18 PRINT 30, NSUB(I),(ILXAB(I,J),J=1,NA),IT1(I),(ILXAT(I,J),J=1,NA),I
     1T2(I)
      DO 19 J=1,NA
      IT1(J)=KSUM(ILXAB,NL,J,NDL)
   19 IT2(J)=KSUM(ILXAT,NL,J,NDL)
      I1=KSUM(IT1,NASCOR,1,1)
      I2=KSUM(IT2,NASCOR,1,1)
      PRINT 31, (IT1(J),J=1,NA),I1,(IT2(J),J=1,NA),I2
   23 PRINT 35
      DO 24 I=1,7
   24 PRINT 36, (ITEM(J),J=I,112,7)
      RETURN
C
   25 FORMAT (15(10(2X,R4,1H=,I4)/))
   41 FORMAT (1H1,10X,'SCORES FOR ',I2,1X,2A6,1X,I3,3A6)
   26 FORMAT (10X,26HQUARTERS OBSERVED-EXPECTED)
   27 FORMAT (23X,17HATTRIBUTE BIAS   ,34X,16HATTRIBUTE TOTAL )
   28 FORMAT (I4,25F5.0)
   29 FORMAT (33X,34HERRORS FOR LISTENERS BY ATTRIBUTES/23X,17HATTRIBUTE
     1 PRESENT,34X,16HATTRIBUTE ABSENT/1X,4H(L#),2(4X,47HVOIC  NASL  SUS
     2T  SIBL  GRAV  COMP  EXPL  TOTAL,1X))
   30 FORMAT (2H (,I2,2H) ,2(8I6,4X))
   31 FORMAT (6H TOTAL,2(8I6,4X))
   32 FORMAT (1H0,37X,28HERRORS FOR QUARTERS BY ITEMS)
   33 FORMAT (7H0ITEM #,3X,14(1H(,I3,1H),1X))
   34 FORMAT (8X,14I6)
   35 FORMAT (1H0,34X,20HERRORS FOR EACH ITEM/10X,10HITEMS 1-28,10X,11HI
     1TEMS 29-56,10X,11HITEMS 57-84,10X,12HITEMS 85-112)
   36 FORMAT (1H ,5X,4(4I4,5X))
      END
```

```
      FUNCTION KSUM (KKK,NN,NRC,ND)
      DIMENSION KKK(ND,1)
      KSUM=0
      K=IABS(NRC)
      N=IABS(NN)
      I=1
      IF (NN) 1,5,2
    1 I=2
    2 IF (NRC) 6,5,3
    3 DO 4 J=1,N
    4 KSUM=KSUM+KKK(J,K)**I
    5 RETURN
    6 DO 7 J=1,N
    7 KSUM=KSUM+KKK(K,J)**I
      RETURN
      END
```

```
      SUBROUTINE FINII
      INTEGER CODE
      COMMON/ALL/CODE(150),SEL(20),NQ,NL,NA,NV,JTTEST,NDL,NASCOR,JPUNCH
     1,NR,JAVE,IBIG,JSPK,IOP(10),ISAVE(200),NATE(20)
      COMMON /MF/ LABLE(13),ITST,NAAME(20),LIIST(20),JPQXR
      COMMON /SCORE/ PAP(7),SAP(7),PAA(7),SAA(7),PAB(7),SAB(7),PAT(7),SA
     1T(7),PV(8),SV(8),PTOT,STOT,RATE(10),SERATE(10),SPO(10, 2,4),SPO2(
     210,2,4)
      IF(LABLE(13).EQ.0)  LABLE(13)=6H
      IF(LABLE(12).EQ.0)  LABLE(12)=6H
      IF(JAVE.NE.1) LABLE(12)=6HMULTI
      PRINT 3,(LABLE(J),J=2,4),(CODE(I),I=1,NQ,4)
      MEOW1=AND(7777B,LABLE(11))
      MEOW2=AND(7777B,LSHIFT(LABLE(11),-12)   )
      MEOW3=AND(7777B,LSHIFT(LABLE(11),-24)       )
      PRINT 4,(LABLE(J),J=7,10),MEOW3,MEOW2,MEOW1,LABLE(12)
      PRINT 5
      PRINT 6, PAP(1),SAP(1),PAA(1),SAA(1),PAB(1),SAB(1),PAT(1),SAT(1)
      PRINT 7, PAP(2),SAP(2),PAA(2),SAA(2),PAB(2),SAB(2),PAT(2),SAT(2)
      PRINT 8, PAP(3),SAP(3),PAA(3),SAA(3),PAB(3),SAB(3),PAT(3),SAT(3)
      PRINT 9, PAP(4),SAP(4),PAA(4),SAA(4),PAB(4),SAB(4),PAT(4),SAT(4)
      PRINT 10, PAP$5),SAP(5),PAA(5),SAA(5),PAB(5),SAB(5),PAT(5),SAT(5)
      PRINT 11, PAP(6),SAP(6),PAA(6),SAA(6),PAB(6),SAB(6),PAT(6),SAT(6)
      PRINT 13,((SPO(9,J,I),SPO2(9,J,I),I=1,4),J=1,2)
      PRINT 12, PAA(7),SAA(7),PAP(7),SAP(7),PAB(7),SAB(7),PAT(7),SAT(7)
      PRINT 14
      INLX=NL
      ISPX=1
      MWOZ=6HNUM OF
      NWOZ=6H LISTE
      JWOZ=6HNERS
      JNIZ=6HNER
      IF(JSPK) 41,42,41
   41 NWOZ=6H SPEAK
      JWOZ=6HERS
      JNIZ=6HER
      INLX=JSPK
      ISPX=NL
   42 CONTINUE
      PRINT 19,INLX,PV(3),SV(3)
      PRINT 20,ISPX,PV(7),SV(7)
      NN=16*NQ*NASCOR*JAVE
      PRINT 21,NN,PV(8),SV(8)
      PRINT 22, PV(4),SV(4)
      JNIZ=JPQXR+1
      IF(JPUNCH.EQ.0) JNIZ=1
      DO 43 I=JNIZ,12
   43 NAAME(I)=6H
      PRINT 17,(NAAME(I),I=1,12),PV(2),SV(2)
      PRINT 15, PV(6),SV(6)
      PRINT 18, PV(1),SV(1)
      PRINT 16,NWOZ,JWOZ,PV(5),SV(5)
      IF (JPUNCH.NE.0) GO TO 1
      PRINT 25
      PRINT 24
      PRINT 23, PTOT
      PRINT 24
      PRINT 27, STOT
      PRINT 24
      PRINT 25
      RETURN
    1 PRINT 29
      PRINT 30, RATE(1),SERATE(1)
      PRINT 31, RATE(2),SERATE(2),PTOT
      PRINT 33, RATE(3),SERATE(3)
      PRINT 32, RATE(4),SERATE(4),STOT
      PRINT 34, RATE(5),SERATE(5)
      IF (NR.GT.5) GO TO 2
      PRINT 26
      PRINT 44
      RETURN
    2 PRINT 28, RATE(6),SERATE(6)
      MEOW1=6H
      MEOW2=6H
      IF(JAVE.LT.2) JWOZ=6H
```

```
      IF(JAVE.LT.2) NAOZ=6H
      IF(JAVE.LT.2) GO TO 8932
      MEOW1=6HAVERAG
      MEOW2=6HED BY
 8932 CONTINUE
      PRINT 44,MEOW1,MEOW2,NWOZ,JWOZ
 44   FORMAT (5X,'(QUALITY RATINGS NOT FOR SCIENTIFIC USE)',19X,4A6)
      RETURN
 C
 3    FORMAT (1H1,4X,32HDIAGNOSTIC RHYME TEST SCORES FOR,3X,3A6,21X,10(R
     14,1X))
 4    FORMAT(5X,22HEXPERIMENTAL CONDITION,2X,4A6,6X,'DATE TESTED ',R2,
     1'/'R2,'/',R2,2X,'LIST NO.',A6/)
 5    FORMAT (5X,9HATTRIBUTE,9X,8HMEAN FOR,3X,4HS.E.,2X,8HMEAN FOR,4X,4H
     1S.E.,2X,8HMEAN FOR,4X,4HS.E.,2X,8HMEAN FOR,4X,4HS.E./23X,9HATTRIBU
     2TE,8X,9HATTRIBUTE,9X,9HATTRIBUTE,9X,9HATTRIBUTE/,23X,7HPRESENT,10X
     3,6HABSENT,12X,5HDIFF./)
 6    FORMAT (5X,7HVOICING,11X,4(F6.1,3X,F6.2,3X)/)
 7    FORMAT (5X,8HNASALITY,10X,4(F6.1,3X,F6.2,3X)/)
 8    FORMAT (5X,10HSUSTENTION,8X,4(F6.1,3X,F6.2,3X)/)
 9    FORMAT (5X,10HSIBILATION,8X,4(F6.1,3X,F6.2,3X)/)
 10   FORMAT (5X,9HGRAVENESS,9X,4(F6.1,3X,F6.2,3X)/)
 11   FORMAT (5X,11HCOMPACTNESS,7X,4(F6.1,3X,F6.2,3X))
 12   FORMAT (5X,13HEXPERIMENTAL ,5X,4(F6.1,3X,F6.2,3X))
 13   FORMAT (8X,15HBACK VS. FRONT ,2X,4(F6.1,3X,F6.2,3X)/8X,15HBACK VS.
     1 MIDDLE,2X,4(F6.1,3X,F6.2,3X))
 14   FORMAT (//57X,13HVOWEL CONTEXT,8X,4HMEAN,8X,4HS.E.)
 15   FORMAT (5X,'REMARKS:1)EXPERIMENTAL ITEMS ARE NOT',20X,4H(OH),11X,
     1F6.1,6X,F6.2)
 16   FORMAT (13X,'2)ALL S.E.',1H','S BASED ON MEANS OF ',2A6,5X,4H(AH),
     1,11X,F6.1,6X,F6.2,/)
 17   FORMAT (5X,10HSPEAKER(S),2X,12A3,8X,4H(OO),11X,F6.1,6X,F6.2)
 18   FORMAT (17X,'INCLUDED IN ANY SUMMARY SCORES.',13X,4H(AW),11X,F6.1,
     16X,F6.2)
 19   FORMAT (5X,'NUMBER OF LISTENERS',I4,33X,4H(EE),11X,F6.1,6X,F6.2)
 20   FORMAT (5X,'NUMBER OF SPEAKERS ',I4,33X,4H(IH),11X,F6.1,6X,F6.2)
 21   FORMAT (5X,'DRT WORDS PRESENTED',I9,28X,4H(EH),11X,F6.1,6X,F6.2)
 22   FORMAT (61X,4H(AT),11X,F6.1,6X,F6.2)
 23   FORMAT (55X,1HX,6X,15HTOTAL DRT SCORE,4X,F6.1,6X,1HX)
 24   FORMAT (55X,1HX,37X,1HX)
 25   FORMAT (55X,39HXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX)
 26   FORMAT (55X,37HXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX)
 27   FORMAT (55X,1HX,6X,14HSTANDARD ERROR,5X,F6.2,6X,1HX)
 28   FORMAT (5X,18HROUGH VS SMOOTH    ,F4.2,3X,F5.2,20X,37HXXXXXXXXXXXXX
     1XXXXXXXXXXXXXXXXXXXXXXXX)
 29   FORMAT (5X,31HQUALITY RATINGS     MEAN     S.E.,19X,39HXXXXXXXXXXXXX
     1XXXXXXXXXXXXXXXXXXXXXXXX   )
 30   FORMAT (5X,18HSOFT VS LOUD       ,F4.2,3XF5.2,20X,1HX,35X,1HX)
 31   FORMAT (5X,18HTREBLE VS BASS     ,F4.2,3XF5.2,20X,1HX,4X,15HTOTAL D
     1RT SCORE,4X,F6.1,6X,1HX)
 32   FORMAT (5X,18HUNPLSNT VS PLSNT   ,F4.2,3XF5.2,20X,1HX,4X,14HSTANDAR
     1D ERROR,5X,F6.2,6X,1HX)
 33   FORMAT (5X,18HUNCLEAR VS CLEAR   ,F4.2,3XF5.2,20X,1HX,35X,1HX)
 34   FORMAT (5X,18HUNNAT. VS NATURAL  ,F4.2,3XF5.2,20X,1HX,35X,1HX)
      END
```

```
      FUNCTION PRBF(DA,DB,FR)
      PRBF=1.0
      IF (DA .LE. 0.0) RETURN
      IF (DB .LE. 0.0) RETURN
      IF (FR .LE. 0.0) RETURN
      IF (FR .LE. 1.0) GO TO 5
      A=DA
      B=DB
      F=FR
      GO TO 10
    5 A=DB
      B=DA
      F=1.0/FR
   10 AA=2.0/(9.0*A)
      BB=2.0/(9.0*B)
      Z=ABS(((1.0-BB)*F**.333333-1.0+AA)/SQRT(BB*F**.666667+AA))
      IF(B.LT.4.0) Z=Z*(1.0+.08*Z**4/B**3)
      PRBF=.5/(1.0+Z*(.196854+Z*(.115194+Z*(.000344+Z*.019527))))**4
      IF(FR.LT.1.0)  PRBF=1.0-PRBF
      RETURN
      END
```

```
      SUBROUTINE FINIS
      INTEGER CODE
      COMMON/ALL/CODE(150),SEL(20),NQ,NL,NA,NV,JTTEST,NDL,NASCOR,JPUNCH
     1,NR,JAVE,IDIG,JSPK,IOP(10),ISAVE(200),NATE(20)
      COMMON /NF/ LABLE(13),ITST,NAAME(20),LIIST(20),JPUKH
      COMMON /SCORE/ PAP(7),SAP(7),PAA(7),SAA(7),PAB(7),SAB(7),PAT(7),SA
     1T(7),PV(8),SV(8),PTOT,STOT,RATE(10),SERATE(10),SPO(10, 2,4),SPO2(
     210,2,4)
      DIMENSION O(10,2,4)
      DO 1 I=1,10
      DO 1 J=1,2
      DO 1 K=1,3
    1 O(I,J,K)=6H
      O(1,1,1)=6HFRICTI
      O(1,1,2)=6HONAL
      O(1,2,1)=6HNONFRI
      O(1,2,2)=6HCTIONA
      O(1,2,3)=6HL
      O(2,1,1)=6HGRAVE
      O(2,2,1)=6HACUTE
      O(3,1,1)=6HVOICED
      O(3,2,1)=6HUNVOIC
      O(3,2,2)=6HED
      O(4,1,1)=6HVOICED
      O(4,2,1)=6HUNVOIC
      O(4,2,2)=6HED
      O(5,1,1)=6HVOICED
      O(5,2,1)=6HUNVOIC
      O(5,2,2)=6HED
      O(6,1,1)=6H   STO
      O(6,1,2)=6HPPED
      O(6,2,1)=6H   UNS
      O(6,2,2)=6HTOPPED
      O(7,1,1)=6HVOICED
      O(7,2,1)=6HUNVOIC
      O(7,2,2)=6HED
      O(8,1,1)=6H   SUS
      O(8,1,2)=6HTAINED
      O(8,2,1)=6H   INT
      O(8,2,2)=6HERRUPT
      O(8,2,3)=6HED
      O(9,1,2)=6HB/M
      O(9,2,2)=6HB/F
      NPZ=NQ+16*NASCOR*JAVE
      MEOW1=AND(7777B,LABLE(11))
      MEOW2=AND(7777B,LSHIFT(LABLE(11),-12))
      MEOW3=AND(7777B,LSHIFT(LABLE(11),-24))
      PRINT 5,(LABLE(I),I=2,4),(LABLE(I),I=7,9),MEOW3,MEOW2,MEOW1
      PRINT 6
      PRINT 7, PAP(1),SAP(1),PAA(1),SAA(1),PAB(1),SAB(1),PAT(1),SAT(1)
      KX=1
      IF (IOP(KX).EQ.0) PRINT 4, (O(KX,1,J),J=1,3),(SPO(KX,1,I),SPO2(KX,
     11,I),I=1,4)
      IF (IOP(KX).EQ.0) PRINT 4, (O(KX,2,J),J=1,3),(SPO(KX,2,I),SPO2(KX,
     12,I),I=1,4)
      PRINT 8, PAP(2),SAP(2),PAA(2),SAA(2),PAB(2),SAB(2),PAT(2),SAT(2)
      KX=KX+1
      IF (IOP(KX).EQ.0) PRINT 4, (O(KX,1,J),J=1,3),(SPO(KX,1,I),SPO2(KX,
     11,I),I=1,4)
      IF (IOP(KX).EQ.0) PRINT 4, (O(KX,2,J),J=1,3),(SPO(KX,2,I),SPO2(KX,
     12,I),I=1,4)
      PRINT 9, PAP(3),SAP(3),PAA(3),SAA(3),PAB(3),SAB(3),PAT(3),SAT(3)
      KX=KX+1
      IF (IOP(KX).EQ.0) PRINT 4, (O(KX,1,J),J=1,3),(SPO(KX,1,I),SPO2(KX,
     11,I),I=1,4)
      IF (IOP(KX).EQ.0) PRINT 4, (O(KX,2,J),J=1,3),(SPO(KX,2,I),SPO2(KX,
     12,I),I=1,4)
      PRINT 10, PAP(4),SAP(4),PAA(4),SAA(4),PAB(4),SAB(4),PAT(4),SAT(4)
      KX=KX+1
      IF (IOP(KX).EQ.0) PRINT 4, (O(KX,1,J),J=1,3),(SPO(KX,1,I),SPO2(KX,
     11,I),I=1,4)
      IF (IOP(KX).EQ.0) PRINT 4, (O(KX,2,J),J=1,3),(SPO(KX,2,I),SPO2(KX,
     12,I),I=1,4)
      PRINT 11, PAP(5),SAP(5),PAA(5),SAA(5),PAB(5),SAB(5),PAT(5),SAT(5)
      KX=KX+1
      IF (IOP(KX).EQ.0) PRINT 4, (O(KX,1,J),J=1,3),(SPO(KX,1,I),SPO2(KX,
     11,I),I=1,4)
      IF (IOP(KX).EQ.0) PRINT 4, (O(KX,2,J),J=1,3),(SPO(KX,2,I),SPO2(KX,
     12,I),I=1,4)
```

```
      KX=KX+1
      IF (IOP(KX).EQ.0) PRINT 4, (O(KX,1,J),J=1,3),(SPO(KA,1,I),SPD2(KA,
     11,I),I=1,4)
      IF (IOP(KX).EQ.C) PRINT 4, (O(KX,2,J),J=1,3),(SPO(KA,2,I),SPD2(KA,
     12,I),I=1,4)
      PRINT 12, PAP(6),SAP(6),PAA(6),SAA(6),PAB(6),SAB(6),PAT(6),SAT(6)
      KX=KX+1
      IF (IOP(KX).EQ.0) PRINT 4, (O(KX,1,J),J=1,3),(SPO(KA,1,I),SPD2(KA,
     11,I),I=1,4)
      IF (IOP(KX).EQ.0) PRINT 4, (O(KX,2,J),J=1,3),(SPO(KA,2,I),SPD2(KA,
     12,I),I=1,4)
      KX=KX+1
      IF (IOP(KX).EQ.0) PRINT 4, (O(KX,1,J),J=1,3),(SPO(KA,1,I),SPD2(KX,
     11,I),I=1,4)
      IF (IOP(KX).EQ.0) PRINT 4, (O(KX,2,J),J=1,3),(SPO(KA,2,I),SPD2(KX,
     12,I),I=1,4)
      KX=KX+1
      IF (IOP(KX).EQ.0) PRINT 4, (O(KX,1,J),J=1,3),(SPO(KA,1,I),SPD2(KX,
     11,I),I=1,4)
      IF (IOP(KX).EQ.0) PRINT 4, (O(KX,2,J),J=1,3),(SPO(KA,2,I),SPD2(KX,
     12,I),I=1,4)
      PRINT 13, PAA(7),SAA(7),PAP(7),SAP(7),PAB(7),SAB(7),PAT(7),SAT(7)
      PRINT 14
      IF(JPQXR.GT.0) GO TO 2
      PRINT 14
   14 FORMAT (//)
      GO TO 3
    2 PRINT 16,(NAAME(J),J=1,JPQXR)
      PRINT 15,(LIIST(J) ,J=1,JPQXR)
   16 FORMAT (5X'SPKR(S)=',1X,20A6)
   15 FORMAT (5X,'LIST(S)=',20A6)
    3
      NA01=6HLISTEN
      NA02=6HERS
      IF(JSPK.EQ.0) GO TO 31
      NA01=6HSPEAKE
      NA02=6HRS
   31 CONTINUE
      PRINT 32
      PRINT 17,NA01,NA02,NL,PTOT
   17 FORMAT(5X,'NUMBER OF ',2A6,I3,37X,1HX,5X,'TOTAL DRT SCORE='F6.1,4X
     1,1HX)
      PRINT 18,NPZ,STOT
   18 FORMAT (5X,'NUMBER OF WORDS PER TEST',I8,30X,1HX,5X,'STANDARD ERRO
     1R=',F6.2,5X,1HX)
      PRINT 32
   32 FORMAT (67X,'XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX')
      RETURN
C
C
    4 FORMAT (5X,2A6,A5,4(F6.1,5X,F6.2,5X))
    5 FORMAT (1H1,4X,'CONTRACTOR: ',3A6,5X,'TEST CONDITION: '3A6,5X,
     1'DATE TESTED  ',R2,'/',R2,'/',R2)
    6 FORMAT (21X,6HPRESNT,7X,4HS.E.,5X,6HABSENT,7X,4HS.E.,5X,4HBIAS,9X,
     14HS.E.,5X,5HTOTAL,6X,4HS.E.)
    7 FORMAT (/4X,7HVOICING,9X4(F6.1,5X,F6.2,5X))
    8 FORMAT (/4X,8HNASALITY,8X,4(F6.1,5X,F6.2,5X))
    9 FORMAT (/4X,10HSUSTENTION,6X,4(F6.1,5X,F6.2,5X))
   10 FORMAT (/4X,10HSIBILATION,6X,4(F6.1,5X,F6.2,5X))
   11 FORMAT (/4X,9HGRAVENESS,7X,4(F6.1,5X,F6.2,5X))
   12 FORMAT (/4X,11HCOMPACTNESS,5X,4(F6.1,5X,F6.2,5X))
   13 FORMAT (/4X,13HEXPERIMENTAL*,3X,4(F6.1,5X,F6.2,5X))
      END
```

```
      FUNCTION SGNF(T,DF)
      SGNF=6H
      IF(DF) 1,2,3
    1 DF=DF*-1
      IF(PRBF(DF,1000.,T/DF).LT. .05)  SGNF=6HP<.05
      IF(PRBF(DF,1000.,T/DF).LT. .01)  SGNF=6HP<.01
      IF(PRBF(DF,1000.,T/DF).LT..001)  SGNF=6HP<.001
      RETURN
    2 SGNF=6HDF = 0
      RETURN
    3 IF(PRBF(1.0,DF,T**2).LT. .05) SGNF=6HP<.05
      IF(PRBF(1.0,DF,T**2).LT. .01) SGNF=6HP<.01
      IF(PRBF(1.0,DF,T**2).LT..001) SGNF=6HP<.001
      RETURN
```

```
      FUNCTION PRBL(DA,DB,F)
      P=PRBF(DA,DB,F)
      IF(P-.10)62,61,61
   61 PRBL=6H
      RETURN
   62 IF(P-.05)64,63,63
   63 PRBL=6HP<.10
      RETURN
   64 IF(P-.01)66,65,65
   65 PRBL=6HP<.05
      RETURN
   66 IF(P-.001)68,67,67
   67 PRBL=6HP<.01
      RETURN
   68 PRBL=6HP<.001
      RETURN
      END
```

```
      FUNCTION ISUB (J)
C     THIS ROUTINE KEEPS A LIST OF SUBJECT NUMBERS USED FOR A LIST
C       AND ALLOWS FOR DIFFERENT SEQUENCES OF SUBJECTS.  IT RETURNS
C       AN ARBITRARY NUMBER WHICH IS CONSISTANT FOR ANY ONE SUBJECT
C       IT ALSO CHECKS FOR BAD KEYPUNCHING OR TOO MANY SUBJECTS
      INTEGER CODE
      COMMON/ALL/CODE(150),SEL(20),NQ,NL,NA,NV,JTTEST,NDL,NASCOR,JPUNCH
     1,NR,JAVE,IBIG,JSPK,IOP(10),ISAVE(200),NATE(20)
      COMMON/MCE/KEY(200,4),JIA,KSAVE(150),NAME(100,2),IANXX(20,4),ISPK
     1EY(10)
      DIMENSION  INAME(20),ISPAR(100)
      IF (J.NE.2HAA) GO TO 4
      DO 1 I=1,100
    1 ISPAR(I)=0
    2 DO 3 I=1,20
    3 INAME(I)=0
      RETURN
    4 CONTINUE
      IF (J.EQ.2HBX) GO TO 2
      IF (J.NE.2HEX) GO TO 7
    5 READ 11, KOP,KSUB,NAM1,NAM2,KPO
      IF (KOP.NE.0) GO TO 6
      NAME(KSUB,1)=NAM1
      NAME(KSUB,2)=NAM2
      IF (KPO.NE.0) RETURN
      GO TO 5
    6 ISPAR(KOP)=KSUB
      IF (KPO.NE.0) RETURN
      GO TO 5
    7 IF(J.LT.-99) GO TO 13
      IF(J.LT.0)  GO TO 16
      IF(ISPAR(J).NE.0) J=ISPAR(J)
      NSUB=NL
      I=0
    8 I=I+1
      IF (I.GT.NSUB) GO TO 10
      IF (J.EQ.INAME(I)) GO TO 9
      IF (INAME(I).NE.0) GO TO 8
      INAME(I)=J
    9 ISUB=I
      RETURN
   10 PRINT 12, NSUB,J
      ISUB=0
      RETURN
   13 IREMU=ABS(J)
      DO 14 K=1,JSPK
   14 IF(NATE(K).EQ.IREMU) GO TO 15
      PRINT 20
   20 FORMAT( '    INELIGIBLE SPEAKER DOWNFIELD')
   15 ISUB=K
      IREMU=K
      RETURN
   16 ISUB=IREMU
      J=IREMU
      RETURN
C
   11 FORMAT (I2,I2,2X,2A6,60X,I2)
   12 FORMAT (10X,21H**********  MORE THAN,1X,I2,1X,11HPRESENT Ss=,I10,1
     10H**********)
      END
```

```
      SUBROUTINE RATING (INC)
      INTEGER CODE
      COMMON /ALL/CODE(150),SEL(20),NQ,NBL,NA,NV,JTTEST,NDL,NASCOR,
     1JPUNCH,NR,JAVE,IB16,JSPK,IOP(10),ISAV(200)
      COMMON /SCORE/ PAP(7),SAP(7),PAA(7),SAA(7),PAB(7),SAB(7),PAT(7),SA
     1T(7),PV(8),SV(8),PTOT,STOT,RATE(10),SERATE(10),SPO(10, 2,4),SPO2(
     210,2,4)
      COMMON /MF/ LABLE(13),ITST,NAAME(20),LIIST(20),JPQXR
      DIMENSION PRATE(10),DRATE(10,20,80)
      INTEGER ANAME,ALIST
      NL=NBL
      IF (INC) 3,5,1
    1 CONTINUE
      DO 2 I=1,80
      DO 2 J=1,NL
      DO 2 K=1,10
    2 DRATE(K,J,I)=0.0
      NQXL=0
      JPQXR=0
      DO 21 I=1,20
      NAAME(I)=0
   21 LIIST(I)=0
      RETURN
    3 ANL=NL
      NQXL=NQXL+1
      I=NQXL
      ANQXL=NQXL
      IMA=1
      IF (INC.LT.-1) IMA=-1
      DO 4 JJ=1,NBL
      READ 11,ANAME,ALIST,IL,(PRATE(K),K=1, 7),LABLE(11),(LABLE(K),K=2,4
     1),(LABLE(K),K=7,9)
   11 FORMAT (4X,A2,A4,I2,7F3.0,A6,1X,3A6,1X,3A6)
      J=ISUB(IL)
      IF(JSPK.GT.0) J=ISUB(-ANAME)
      DO 4 K=1,NR
    4 DRATE(K,J,I)=DRATE(K,J,I)+PRATE(K)*IMA
      DO 41 JJ=1,JPQXR
      IF(NAAME(JJ).EQ.ANAME) GO TO 44
   41 CONTINUE
      JPQXR=JPQXR+1
      NAAME(JPQXR)=ANAME
      LIIST(JPQXR)=ALIST
   44 RETURN
    5 DO 8 K=1,NR
      TEMP2=0.0
      DO 7 J=1,NL
      TEMP=0.0
      DO 6 I=1,NQXL
    6 TEMP=DRATE(K,J,I)+TEMP
      DRATE(K,J,1)=TEMP/ANQXL
    7 TEMP2=TEMP/ANQXL+TEMP2
      RATE(K)=TEMP2/ANL
    8 CONTINUE
      DO 10 K=1,NR
      TEMP=0.0
      DO 9 J=1,NL
    9 TEMP=((DRATE(K,J,1)-RATE(K))**2)+TEMP
   10 SERATE(K)=SQRT(TEMP/(NL-1))/SQRT(NL)
      IF(JAVE.GT.1) GO TO 12
      LABLE(12)=ALIST
      LABLE(13)=ANAME
   12 CONTINUE
      IF(JTTEST.NE.0) GO TO 13
      RATE(1)=8.-RATE(1)
      RATE(2)=8.-RATE(2)
      RATE(4)=8.-RATE(4)
      RATE(6)=8.-RATE(6)
   13 CONTINUE
      RETURN
      END
```

```
      SUBROUTINE STDPC
      INTEGER CODE
      COMMON/ALL/CODE(150),SEL(20),NXQ,NL,NA,NV,JTTEST,NDL,NASCOR,JPUNCH
     1,NR,JAVE,IHIG,JSPK,IOP(10),ISAVE(200),NATE(20)
      COMMON /SCORE/ PAP(7),SAP(7),PAA(7),SAA(7),PAB(7),SAB(7),PAT(7),SA
     1T(7),PV(8),SV(8),PTOT,STOT,RATE(10),SERATE(10),SPO(10, 2,4),SPO2(
     210,2,4)
      COMMON /ERR/ NSUB(20),ILXAP(20,7),ILXAA(20,7),ILXAB(20,7),ILXAT(20
     1,7),ILXQ(20,150),ITEM(112),ISPLT(10,20,2,4),ILXV(20,8)
      CALL RATING (0)
      T=0.0
      IF (JTTEST.NE.0) T=100.00
      NQ=NXQ*JAVE
      AA=NQ*NL
      IF(JSPK.GT.0) GO TO 8
    9 UNBS=NL*(NL-1)
      DO 1 I=1,NA
      SX=KSUM(ILXAP,NL,I,NDL)
      SX2=KSUM(ILXAP,-NL,I,NDL)
      SAP(I)=SQRT(625.*(SX2-SX**2/NL)/UNBS)/NQ
      PAP(I)=((AA*8.-(2*SX))/(AA*8.))*100.-T
      SX=KSUM(ILXAA,NL,I,NDL)
      SX2=KSUM(ILXAA,-NL,I,NDL)
      SAA(I)=SQRT(625.*(SX2-SX**2/NL)/UNBS)/NQ
      PAA(I)=((AA*8.-(2*SX))/(AA*8.))*100.-T
      SX=KSUM(ILXAB,NL,I,NDL)
      SX2=KSUM(ILXAB,-NL,I,NDL)
      SAB(I)=SQRT(625.*(SX2-SX**2/NL)/UNBS)/NQ
      PAB(I)=PAP(I)-PAA(I)
      SX=KSUM(ILXAT,NL,I,NDL)
      SX2=KSUM(ILXAT,-NL,I,NDL)
      SAT(I)=SQRT(156.25*(SX2-SX**2/NL)/UNBS)/NQ
    1 PAT(I)=(PAP(I)+PAA(I))/2.
      GSX=0.
      GSX2=0.
      NHQ=NXQ
      IF(JSPK.EQ.0) NHQ=NQ
      DO 2 I=1,NL
      SX=0.0
      SX2=0.0
      SZX=0.0
      SZX2=0.0
      XB=96.0
      IF(JSPK.NE.0) XB=JSPK*96.0
      DO 10 KQ=1,NHQ
      SX=SX+ILXQ(I,KQ)
      SX2=SX2+ILXQ(I,KQ)**2
      SZX=SZX+((XB-2.0*ILXQ(I,KQ))/XB)*100.0
   10 SZX2=SZX2+(((XB-2.0*ILXQ(I,KQ))/XB)*100.0)**2
      BSX=((NQ*96.0-(2*SX))/(NQ*96.))*100.-T
      GSX2=GSX2+BSX**2
      GSX=GSX+BSX
    2 SEL(I)=SQRT(ABS((SZX2-SZX**2/NHQ)/(NHQ*(NHQ-1))))      )
      STOT=SQRT(ABS((GSX2-GSX**2/NL)/(NL*(NL-1))))
      PTOT=GSX/NL
      DO 5 K=1,9
      DO 5 J=1,2
      DO 4 I=1,4
      SPO(K,J,I)=0.0
      SPO2(K,J,I)=0.0
      DO 3 L=1,NL
      SPO(K,J,I)=SPO(K,J,I)+ISPLT(K,L,J,I)
    3 SPO2(K,J,I)=SPO2(K,J,I)+ISPLT(K,L,J,I)**2
      SPO2(K,J,I)=2.*SQRT(625.*(SPO2(K,J,I)-SPO(K,J,I)**2/NL)/UNBS)/NQ
      IF(I.EQ.4) SPO2(K,J,I)=SPO2(K,J,I)/2.
    4 SPO(K,J,I)=(AA*4.-2.*SPO(K,J,I))/(AA*4.)*100.-T
      SPO(K,J,3)=SPO(K,J,1)-SPO(K,J,2)
      SPO(K,J,4)=(SPO(K,J,1)+SPO(K,J,2))/2.0
    5 CONTINUE
      DO 7 I=1,8
      SX=0.0
      SX2=0.0
      DO 6 J=1,NL
      SX=SX+ILXV(J,I)
    6 SX2=SX2+ILXV(J,I)**2
      SV(I)=SQRT(278.5554*(SX2-SX**2/NL)/UNBS)/NQ
    7 PV(I)=((AA*12.0-(2*SX))/(AA*12.0))*100.0-T
      RETURN
    8 ITE=NL
      NQ=NXQ*NL*(JAVE/JSPK)
      NL=JSPK
      JSPK=ITE
      GO TO 9
      END
```

```
      SUBROUTINE TTST
      INTEGER CODE
      COMMON/ALL/CODE(150),SEL(20),NQ,NL,NA,NV,JTTEST,NDL,NASCOR,JPUNCH
     1,NR,JAVE,IBIG,JSPK,IOP(10),ISAVE(200),NATL(20)
      COMMON /SCORE/ PAP(7),SAP(7),PAA(7),SAA(7),PAB(7),SAB(7),PAT(7),SA
     1T(7),PV(8),SV(8),PTOT,STOT,RATE(10),SERATE(10),SPO(10, 2,4),SPO2(
     210,2,4)
      DIMENSION SIG(5,8)
      DF=NL-1.0
      NDF=NL-1
      DO 1 I=1,NA
      PAP(I)=ABS(PAP(I)/SAP(I))
      PAA(I)=ABS(PAA(I)/SAA(I))
      PAB(I)=ABS(PAB(I)/SAB(I))
    1 PAT(I)=ABS(PAT(I)/SAT(I))
      DO 2 I=1,NV
    2 PV(I)=ABS(PV(I)/SV(I))
      PTOT=ABS(PTOT/STOT)
      DO 3 I=1,NA
      SIG(1,I)=SGNF(PAP(I),DF)
      SIG(2,I)=SGNF(PAA(I),DF)
      SIG(3,I)=SGNF(PAB(I),DF)
    3 SIG(4,I)=SGNF(PAT(I),DF)
      SIGTOT=SGNF(PTOT,DF)
      PRINT 15, (CODE(I),I=1,NQ,4)
      PRINT 16
      PRINT 17, PAP(1),SIG(1,1),PAA(1),SIG(2,1),PAB(1),SIG(3,1),PAT(1),S
     1IG(4,1)
      PRINT 18, PAP(2),SIG(1,2),PAA(2),SIG(2,2),PAB(2),SIG(3,2),PAT(2),S
     1IG(4,2)
      PRINT 19, PAP(3),SIG(1,3),PAA(3),SIG(2,3),PAB(3),SIG(3,3),PAT(3),S
     1IG(4,3)
      PRINT 20, PAP(4),SIG(1,4),PAA(4),SIG(2,4),PAB(4),SIG(3,4),PAT(4),S
     1IG(4,4)
      PRINT 21, PAP(5),SIG(1,5),PAA(5),SIG(2,5),PAB(5),SIG(3,5),PAT(5),S
     1IG(4,5)
      PRINT 22, PAP(6),SIG(1,6),PAA(6),SIG(2,6),PAB(6),SIG(3,6),PAT(6),S
     1IG(4,6)
      PRINT 23, PAP(7),SIG(1,7),PAA(7),SIG(2,7),PAB(7),SIG(3,7),PAT(7),S
     1IG(4,7)
      IF (NR.EQ.0) GO TO 7
      DO 4 K=1,NR
      RATE(K)=ABS(RATE(K)/SERATE(K))
      SERATE(K)=SGNF(RATE(K),DF)
    4 CONTINUE
      PRINT 9
      PRINT 10, RATE(1),SERATE(1)
      PRINT 11, RATE(2),SERATE(2)
      PRINT 12, RATE(3),SERATE(3),PTOT,SIGTOT
      PRINT 13, RATE(4),SERATE(4)
      PRINT 14, RATE(5),SERATE(5)
      IF (NR-5) 6,6,5
    5 PRINT 8, RATE(6),SERATE(6)
    6 RETURN
    7 PRINT 24, PTOT,SIGTOT
      RETURN
C
C
    8 FORMAT (5X,15HSMOOTHNESS      ,F6.2,2X,A6)
    9 FORMAT (5X,30HQUALITY RATINGS    T          )
   10 FORMAT (5X,15HLOUDNESS        ,F6.2,2X,A6,16X,45HXXXXXXXXXXXXXXXXXX
     1XXXXXXXXXXXXXXXXXXXXXXXXXXX)
   11 FORMAT (5X,15HBASSINESS       ,F6.2,2X,A6,16X,1HX,43X,1HX)
   12 FORMAT (5X,15HPLEASENTNESS    ,F6.2,2X,A6,16X,1HX,4X,15HTOTAL DRT S
     1CORE,4X,F6.2,4X,A6,4X,1HX)
   13 FORMAT (5X,15HCLEARITY        ,F6.2,2X,A6,16X,1HX,43X,1HX)
   14 FORMAT (5X,15HNATURALNESS     ,F6.2,2X,A6,16X,45HXXXXXXXXXXXXXXXXXX
     1XXXXXXXXXXXXXXXXXXXXXXXXXXX)
   15 FORMAT (1H1,7X,84HT TEST--A TEST FOR SIGNIFICANCE OF DIFFERENCES B
     1ETWEEN DIAGNOSTIC RHYME TEST SCORES.//8X,24HEXPERIMENTAL CONDITION
     2S:,10(2X,A4)//)
   16 FORMAT (1H0,4X,9HATTRIBUTE,7X,7HPRESENT,12X,6HABSENT,14X,4HBIAS,15
     1X,4HMEAN)
   17 FORMAT (1H0,4X,12HVOICING     ,4X,4(F6.2,2X,A6,5X))
   18 FORMAT (1H0,4X,12HNASALITY    ,4X,4(F6.2,2X,A6,5X))
   19 FORMAT (1H0,4X,12HSUSTENTION  ,4X,4(F6.2,2X,A6,5X))
   20 FORMAT (1H0,4X,12HSIBILATION  ,4X,4(F6.2,2X,A6,5X))
   21 FORMAT (1H0,4X,12HGRAVENESS   ,4X,4(F6.2,2X,A6,5X))
   22 FORMAT (1H0,4X,12HCOMPACTNESS ,4X,4(F6.2,2X,A6,5X))
   23 FORMAT (1H0,4X,12HEXPERIMENTAL,4X,4(F6.2,2X,A6,5X))
   24 FORMAT (////,55X,15HTOTAL DRT SCORE,4X,F6.2,4X,A6)
      END
```

```
      SUBROUTINE PUNCH
      INTEGER CODE
      COMMON/ALL/CODE(150),SEL(20),NG,NL,NA,NV,JTTEST,NDL,NASCOR,JPUNCH
     1,NR,JATE,IBIG,JSPK,IUP(10),ISAV(200),NATE(20)
      COMMON /MF/ LABLE(13),ITST
      COMMON /SCORE/ PAP(7),SAP(7),PAA(7),SAA(7),PAB(7),SAB(7),PAT(7),SA
     1T(7),PV(8),SV(6),PTOT,STOT,RATE(10),SERATE(10),SPO(10, 2,4),SPO2(
     210,2,4)
      PUNCH 5, CODE(4),NG,NL,NA,NV,NR,LABLE(12),LABLE(13)
      PUNCH 2, (PAB(J),J=1,7),(PAT(I),I=1,7)
      PUNCH 3, (SAB(J),J=1,7),(SAT(I),I=1,7)
      PUNCH 3, (SAP(J),J=1,7),(SAA(I),I=1,7)
      PUNCH 3, (RATE(I),I=1,NR),(SERATE(J),J=1,NR),PTOT,STOT
      DO 1 I=1,9,2
    1 PUNCH 4, ((SPO(I,J,K),SPO2(I,J,K),J=1,2),K=3,4),((SPO(I+1,J,K),SPO
     12(I+1,J,K),J=1,2),K=3,4)
      RETURN
C
C
    2 FORMAT (7F6.1,7F5.1)
    3 FORMAT (14F5.2)
    4 FORMAT (8(F5.1,F5.2))
    5 FORMAT (A4,5I2,2A6)
      END
```

```
      FUNCTION SUMX2(M,N)
      COMMON /ADD/ SUM(20,2,2,2,2,2),K(10)
      DIMENSION SM(20,2,2,2,2,2),N(10)
      K1=K(1)
      K2=K(2)
      K3=K(3)
      K4=K(4)
      K5=K(5)
      K6=K(6)
      DO 99 I1=1,K1
      DO 99 I2=1,K2
      DO 99 I3=1,K3
      DO 99 I4=1,K4
      DO 99 I5=1,K5
      DO 99 I6=1,K6
   99 SM(I1,I2,I3,I4,I5,I6)=SUM(I1,I2,I3,I4,I5,I6)
      IF(M.EQ.0) GO TO 101
      DO 100 I=1,M
      NN=N(I)
      N(I)=0
      GO TO (1,2,3,4,5,6),NN
    1 DO 11 I2=1,K2
      DO 11 I3=1,K3
      DO 11 I4=1,K4
      DO 11 I5=1,K5
      DO 11 I6=1,K6
      SX=0.
      DO 10 I1=1,K1
      SX=SX+SM(I1,I2,I3,I4,I5,I6)
   10 SM(I1,I2,I3,I4,I5,I6)=0.
   11 SM(1,I2,I3,I4,I5,I6)=SX
      GO TO 100
    2 DO 21 I1=1,K1
      DO 21 I3=1,K3
      DO 21 I4=1,K4
      DO 21 I5=1,K5
      DO 21 I6=1,K6
      SX=0.
      DO 20 I2=1,K2
      SX=SX+SM(I1,I2,I3,I4,I5,I6)
   20 SM(I1,I2,I3,I4,I5,I6)=0.
   21 SM(I1,1,I3,I4,I5,I6)=SX
      GO TO 100
    3 DO 31 I1=1,K1
      DO 31 I2=1,K2
      DO 31 I4=1,K4
      DO 31 I5=1,K5
      DO 31 I6=1,K6
      SX=0.
      DO 30 I3=1,K3
      SX=SX+SM(I1,I2,I3,I4,I5,I6)
   30 SM(I1,I2,I3,I4,I5,I6)=0.
   31 SM(I1,I2,1,I4,I5,I6)=SX
      GO TO 100
```

```
      4      DO 41 I1=1,K1
             DO 41 I2=1,K2
             DO 41 I3=1,K3
             DO 41 I5=1,K5
             DO 41 I6=1,K6
             SX=0.
             DO 40 I4=1,K4
             SX=SX+SM(I1,I2,I3,I4,I5,I6)
      40     SM(I1,I2,I3,I4,I5,I6)=0.
      41     SM(I1,I2,I3,1,I5,I6)=SX
             GO TO 100
      5      DO 51 I1=1,K1
             DO 51 I2=1,K2
             DO 51 I3=1,K3
             DO 51 I4=1,K4
             DO 51 I6=1,K6
             SX=0.
             DO 50 I5=1,K5
             SX=SX+SM(I1,I2,I3,I4,I5,I6)
      50     SM(I1,I2,I3,I4,I5,I6)=0.
      51     SM(I1,I2,I3,I4,1,I6)=SX
             GO TO 100
      6      DO 61 I1=1,K1
             DO 61 I2=1,K2
             DO 61 I3=1,K3
             DO 61 I4=1,K4
             DO 61 I5=1,K5
             SX=0.
             DO 60 I6=1,K6
             SX=SX+SM(I1,I2,I3,I4,I5,I6)
      60     SM(I1,I2,I3,I4,I5,I6)=0.
      61     SM(I1,I2,I3,I4,I5,1)=SX
      100    CONTINUE
      101    SUMX2=0.
             DO 111 I1=1,K1
             DO 111 I2=1,K2
             DO 111 I3=1,K3
             DO 111 I4=1,K4
             DO 111 I5=1,K5
             DO 111 I6=1,K6
      111    SUMX2=SUMX2+SM(I1,I2,I3,I4,I5,I6)**2
             RETURN
             END
```

APPENDIX II

SPECIMEN DRT IV ANSWER BOOKLET

A

BOB - GOB
DAUNT - TAUNT
MOOT - BOOT
SHEET - CHEAT
GAB - JAB
TOT - POT
BOAST - GHOST
RIP - LIP
SAID - ZED
GNAW - DAW
SHOES - CHOOSE
KEEP - CHEEP
DANK - BANK
DOT - GOT
ROAD - LOAD
TINT - DINT
DECK - NECK
TONG - THONG
CHEW - COO
REED - WEED
SAG - SHAG
LOT - ROT
FOAL - VOLE
DIP - NIP
FENCE - PENCE
THAW - SAW
POOL - TOOL
YIELD - WIELD
LAP - RAP

COOT - TOOT
POND - BOND
BONE - MOAN
BILL - VILL
GUEST - JEST
FOUGHT - THOUGHT
POOP - COOP
LEAP - REAP
FAST - VAST
KNOCK - DOCK
DOZE - THOSE
SING - THING
NET - MET
CAUGHT - TAUGHT
LEWD - RUDE
BEAN - PEEN
MAD - BAD
BOX - VOX
JOE - GO
DID - BID
WREN - YEN
LAW - RAW
ZOO - SUE
NEED - DEED
THAN - DAN
CHOP - COP
FORE - THOR
FIT - HIT
LEST - REST

NAME ____________________

DATE ________________

B

PEST - TEST	FAN - PAN
FAULT - VAULT	CHOCK - JOCK
NEWS - DUES	NOTE - DOTE
VEE - BEE	THICK - TICK
THANK - SANK	CHAIR - CARE
WAD - ROD	DONG - BONG
SO - SHOW	RUE - YOU
RID - LID	REEK - LEAK
DENSE - TENSE	GAFF - CALF
BOSS - MOSS	MOM - BOMB
FOO - POOH	DOUGH - THOUGH
THEE - ZEE	GILT - JILT
FAD - THAD	TENT - PENT
FOP - HOP	YAWL - WALL
ROW - LOW	ROOT - LOOT
GIN - CHIN	FEEL - VEAL
BEND - MEND	NAB - DAB
SHAW - CHAW	BON - VON
GOOSE - JUICE	THOLE - SOLE
PEAK - TEAK	THIN - FIN
GAT - BAT	KEG - PEG
ROCK - LOCK	WRONG - LONG
COAT - GOAT	TUNE - DUNE
BIT - MIT	BEAT - MEAT
DEN - THEN	CHAD - SHAD
JAWS - GAUZE	JOT - GOT
MOON - NOON	BOWL - DOLE
TEA - KEY	GILL - DILL
RAMP - LAMP	REND - LEND

GOB - BOB	COOT - TOOT
TAUNT - DAUNT	POND - BOND
MOOT - BOOT	BONE - MOAN
SHEET - CHEAT	BILL - VILL
GAB - JAB	GUEST - JEST
TOT - POT	THOUGHT - FOUGHT
BOAST - GHOST	POOP - COOP
RIP - LIP	REAP - LEAP
SAID - ZED	VAST - FAST
DAW - GNAW	KNOCK - DOCK
SHOES - CHOOSE	DOZE - THOSE
KEEP - CHEEP	SING - THING
DANK - BANK	NET - MET
DOT - GOT	CAUGHT - TAUGHT
ROAD - LOAD	LEWD - RUDE
TINT - DINT	PEEN - BEAN
DECK - NECK	MAD - BAD
THONG - TONG	BOX - VOX
CHEW - COO	JOE - GO
WEED - REED	DID - BID
SAG - SHAG	WREN - YEN
LOT - ROT	LAW - RAW
FOAL - VOLE	SUE - ZOO
DIP - NIP	DEED - NEED
FENCE - PENCE	DAN - THAN
SAW - THAW	CHOP - COP
POOL - TOOL	FORE - THOR
WIELD - YIELD	FIT - HIT
LAP - RAP	REST - LEST

TEST - PEST	FAN - PAN
VAULT - FAULT	CHOCK - JOCK
NEWS - DUES	NOTE - DOTE
VEE - BEE	THICK - TICK
THANK - SANK	CHAIR - CARE
WAD - ROD	BONG - DONG
SO - SHOW	RUE - YOU
RID - LID	LEAK - REEK
DENSE - TENSE	CALF - GAFF
MOSS - BOSS	MOM - BOMB
FOO - POOH	DOUGH - THOUGH
THEE - ZEE	GILT - JILT
FAD - THAD	TENT - PENT
FOP - HOP	YAWL - WALL
ROW - LOW	ROOT - LOOT
GIN - CHIN	VEAL - FEEL
BEND - MEND	NAB - DAB
CHAW - SHAW	BON - VON
GOOSE - JUICE	THOLE - SOLE
TEAK - PEAK	THIN - FIN
GAT - BAT	KEG - PEG
ROCK - LOCK	WRONG - LONG
COAT - GOAT	DUNE - TUNE
BIT - MIT	MEAT - BEAT
DEN - THEN	SHAD - CHAD
GAUZE - JAWS	JOT - GOT
MOON - NOON	BOWL - DOLE
KEY - TEA	GILL - DILL
RAMP - LAMP	LEND - REND

E

GOB - BOB
TAUNT - DAUNT
MOOT - BOOT
SHEET - CHEAT
JAB - GAB
POT - TOT
BOAST - GHOST
RIP - LIP
SAID - ZED
DAW - GNAW
SHOES - CHOOSE
CHEEP - KEEP
BANK - DANK
GOT - DOT
LOAD - ROAD
DINT - TINT
DECK - NECK
THONG - TONG
CHEW - COO
WEED - REED
SAG - SHAG
LOT - ROT
VOLE - FOAL
NIP - DIP
FENCE - PENCE
SAW - THAW
POOL - TOOL
WIELD - YIELD
LAP - RAP

TOOT - COOT
BOND - POND
MOAN - BONE
VILL - BILL
GUEST - JEST
THOUGHT - FOUGHT
COOP - POOP
REAP - LEAP
VAST - FAST
DOCK - KNOCK
DOZE - THOSE
THING - SING
NET - MET
TAUGHT - CAUGHT
LEWD - RUDE
PEEN - BEAN
MAD - BAD
BOX - VOX
GO - JOE
DID - BID
WREN - YEN
LAW - RAW
SUE - ZOO
DEED - NEED
DAN - THAN
COP - CHOP
FORE - THOR
FIT - HIT
REST - LEST

TEST - PEST	FAN - PAN
VAULT - FAULT	CHOCK - JOCK
NEWS - DUES	NOTE - DOTE
VEE - BEE	TICK - THICK
THANK - SANK	CHAIR - CARE
WAD - ROD	BONG - DONG
SO - SHOW	YOU - RUE
RID - LID	LEAK - REEK
TENSE - DENSE	CALF - GAFF
MOSS - BOSS	BOMB - MOM
POOH - FOO	DOUGH - THOUGH
THEE - ZEE	GILT - JILT
FAD - THAD	TENT - PENT
FOP - HOP	YAWL - WALL
LOW - ROW	LOOT - ROOT
GIN - CHIN	VEAL - FEEL
MEND - BEND	DAB - NAB
CHAW - SHAW	VON - BON
GOOSE - JUICE	SOLE - THOLE
TEAK - PEAK	FIN - THIN
BAT - GAT	KEG - PEG
LOCK - ROCK	LONG - WRONG
GOAT - COAT	DUNE - TUNE
BIT - MIT	MEAT - BEAT
DEN - THEN	SHAD - CHAD
GAUZE - JAWS	JOT - GOT
NOON - MOON	DOLE - BOWL
KEY - TEA	DILL - GILL
LAMP - RAMP	LEND - REND

BOB - GOB	TOOT - COOT
DAUNT - TAUNT	BOND - POND
MOOT - BOOT	MOAN - BONE
SHEET - CHEAT	VILL - BILL
JAB - GAB	GUEST - JEST
POT - TOT	FOUGHT - THOUGHT
BOAST - GHOST	COOP - POOP
RIP - LIP	LEAP - REAP
SAID - ZED	FAST - VAST
GNAW - DAW	DOCK - KNOCK
SHOES - CHOOSE	DOZE - THOSE
CHEEP - KEEP	THING - SING
BANK - DANK	NET - MET
GOT - DOT	TAUGHT - CAUGHT
LOAD - ROAD	LEWD - RUDE
DINT - TINT	BEAN - PEEN
DECK - NECK	MAD - BAD
TONG - THONG	BOX - VOX
CHEW - COO	GO - JOE
REED - WEED	DID - BID
SAG - SHAG	WREN - YEN
LOT - ROT	LAW - RAW
VOLE - FOAL	ZOO - SUE
NIP - DIP	NEED - DEED
FENCE - PENCE	THAN - DAN
THAW - SAW	COP - CHOP
POOL - TOOL	FORE - THOR
YIELD - WIELD	FIT - HIT
LAP - RAP	LEST - REST

DRT IV-(1)

H

PEST - TEST	FAN - PAN
FAULT - VAULT	CHOCK - JOCK
NEWS - DUES	NOTE - DOTE
VEE - BEE	TICK - THICK
THANK - SANK	CHAIR - CARE
WAD - ROD	DONG - BONG
SO - SHOW	YOU - RUE
RID - LID	REEK - LEAK
TENSE - DENSE	GAFF - CALF
BOSS - MOSS	BOMB - MOM
POOH - FOO	DOUGH - THOUGH
THEE - ZEE	GILT - JILT
FAD - THAD	TENT - PENT
FOP - HOP	YAWL - WALL
LOW - ROW	LOOT - ROOT
GIN - CHIN	FEEL - VEAL
MEND - BEND	DAB - NAB
SHAW - CHAW	VON - BON
GOOSE - JUICE	SOLE - THOLE
PEAK - TEAK	FIN - THIN
BAT - GAT	KEG - PEG
LOCK - ROCK	LONG - WRONG
GOAT - COAT	TUNE - DUNE
BIT - MIT	BEAT - MEAT
DEN - THEN	CHAD - SHAD
JAWS - GAUZE	JOT - GOT
NOON - MOON	DOLE - BOWL
TEA - KEY	DILL - GILL
LAMP - RAMP	REND - LEND

APPENDIX III

SPECIMEN OUTPUT OF DRT IV COMPUTER SCORING PROGRAM

DIAGNOSTIC RHYME TEST SCORES FOR DRT SCORING DEMONSTRATION
EXPERIMENTAL CONDITION NOISE MASKED SPEECH S/N RATIO: +12DB LIST NO. MULTIPLE

ATTRIBUTE	MEAN FOR ATTRIBUTE PRESENT	S.E.	MEAN FOR ATTRIBUTE ABSENT	S.E.	MEAN FOR ATTRIBUTE DIFF.	S.E.	MEAN FOR ATTRIBUTE	S.E. **
VOICING	96.1	1.09	92.4	2.04	3.7	2.06	94.2	1.27
NASALITY	98.3	.59	99.0	.31	-.7	.49	98.6	.41
SUSTENTION	82.7	5.10	89.3	1.80	-6.5	4.75	86.0	3.07
SIBILATION	97.1	.73	98.6	.20	-1.6	.64	97.9	.33
GRAVENESS	81.8	2.94	89.1	2.59	-7.3	4.28	85.5	1.76
COMPACTNESS	98.0	.72	97.5	.99	.5	1.40	97.8	.52
BACK VS. FRONT	97.9	.33	97.3	1.67	.7	1.66	97.6	.66
BACK VS. MIDDLE	98.0	1.21	97.8	.77	.3	1.78	97.9	.40
EXPERIMENTAL*	98.4	.67	99.0	.33	.7	1.08	98.7	.37

VOWEL CONTEXT	MEAN	S.E.
(EE)	92.6	.99
(IH)	89.6	1.34
(EH)	96.2	.69
(AT)	92.5	.39
(OO)	95.4	.47
(OH)	95.4	1.10
(AW)	93.9	.71
(AH)	91.1	1.99

NUM OF SPEAKERS 6
DRT WORDS PER LISTENER 4608

SPEAKER(S) MULTIPLE
REMARKS: * NOT INCLUDED IN CALCULATION OF THE OTHER SCORES.

QUALITY RATINGS***	MEAN	S.E.
SOFT VS LOUD	3.76	.88
TREBLE VS BASS	4.10	.65
UNCLEAR VS CLEAR	5.34	1.17
UNPLSNT VS PLSNT	4.33	.80
UNNAT. VS NATURAL	4.70	1.03
ROUGH VS SMOOTH	4.45	.78

TOTAL DRT SCORE 93.3

STANDARD ERROR .40**

** ALL STANDARD ERRORS BASED ON SPEAKER MEANS
*** NOT FOR SCIENTIFIC USE

MOST DIFFICULT ITEMS

FOR THE PURPOSES OF FURTHER RESEARCH DESIGNED TO IMPROVE YOUR SYSTEM OR DEVICE, YOU WILL FIND IT ADVANTAGEOUS TO GIVE SPECIAL ATTENTION TO THE DISTINGUISHABILITY OF THE FOLLOWING WORD PAIRS.•

WORD PAIRS	ADJ. PERCENT CORRECT
103:FIN/THIN ••	31.3
68:FAD/THAD ••	50.0
45:VOX/BOX ••	62.0
57:VAULT/FAULT ••	65.6
101:VON/BON ••	66.1
59:VEE/BEE ••	66.1
31:VILL/BILL ••	70.3
33:FOUGT/THOUGT ••	75.5
22:VOLE/FOAL	83.9
10:SHOES/CHOOSE	83.9

•• THE CONTRASTS: FAD-THAD, FIN-THIN, FOUGHT-THOUGHT, VON-BON, VOX-BOX, VEE-BEE, VILL-BILL, VAULT-FAULT ARE GENERALLY AMOUNG THE MOST DIFFICULT TO DISTINGUISH. THEIR PRESENCE ON THE FOREGOING LIST DOES NOT, THEREFORE, REFLECT UNIQUELY UPON THE PERFORMANCE OF YOUR SYSTEM OR DEVICE.

DETAILED DIAGNOSTIC ANALYSIS

	PRESNT	S.E.	ABSENT	S.E.	BIAS	S.E.	TOTAL	S.E.
VOICING	96.1	1.09	92.4	2.04	3.7	2.06	94.2	1.27
FRICTIONAL	93.1	1.93	85.4	4.20	7.7	4.47	89.3	2.38
NONFRICTIONAL	99.1	.51	99.3	.31	-.3	.66	99.2	.27
NASALITY	98.3	.59	99.0	.31	-.7	.49	98.6	.41
GRAVE	98.4	.64	98.4	.53	.0	.49	98.4	.53
ACUTE	98.2	.72	99.5	.16	-1.3	.66	98.8	.40
SUSTENTION	82.7	5.18	89.3	1.80	-6.5	4.75	86.0	3.07
VOICED	74.1	10.87	85.2	4.70	-11.1	11.96	79.6	5.86
UNVOICED	91.4	3.02	93.4	2.48	-2.0	4.63	92.4	1.51
SIBILATION	97.1	.73	98.6	.20	-1.6	.84	97.9	.33
VOICED	96.0	1.33	98.7	.26	-2.7	1.52	97.3	.58
UNVOICED	98.2	.39	99.0	.39	-.8	.61	98.6	.24
GRAVENESS	81.8	2.94	89.1	2.59	-7.3	4.20	85.5	1.76
VOICED	95.3	1.41	96.6	1.35	-1.3	1.72	96.0	1.08
UNVOICED	68.4	4.86	81.6	5.29	-13.3	7.54	75.0	3.40
PLOSIVE	93.1	2.29	91.7	4.44	1.4	5.09	92.4	2.45
NONPLOSIVE	70.6	4.27	86.6	5.16	-16.0	7.69	78.6	2.76
COMPACTNESS	98.0	.72	97.5	.99	.5	1.40	97.8	.52
VOICED	98.8	.27	99.2	.29	-.4	.27	99.0	.24
UNVOICED	97.1	1.51	95.8	1.89	1.3	2.72	96.5	1.03
SUSTAINED	98.8	.48	99.1	.37	-.3	.48	99.0	.36
INTERRUPTED	97.1	1.35	96.0	1.64	1.2	2.63	96.5	.73
B/M	97.9	.33	97.3	1.67	.7	1.06	97.6	.88
B/F	98.0	1.21	97.8	.77	.3	1.78	97.9	.40
EXPERIMENTAL.	98.4	.87	99.0	.33	.7	1.08	98.7	.37

LIST NO.= MULTIPLE
SPKRS NAME MULTIPLE
6 SPEAKERS
4608 WORDS
SN49UZS1SNS00ZS0

TOTAL DRT SCORE 93.3

STANDARD ERROR .40

ERROR ANALYSIS

SCORES FOR 6 SPEAKERS ON 24 HALF TESTS

SN49=	33	NS06=	33	NZ51=	35	ZN16=	58	SN50=	70	NS05=	68	NZ50=	51	YU6A=	42	SN51=	51	NS04=			
NZ49=	45	ZN18=	39	SN52=	60	NS03=	70	NZ54=	81	ZN13=	41	SN53=	60	NS02=	58	NZ53=	59	ZN14=			
SN54=	71	NS01=	49	NZ52=	48	ZN15=	49														

1 R DAVIS	94.9	2.7	3.1	68	89	0	0	-7.2	7.2	.0	.0
2 J EDDINS	93.0	2.1	1.4	100	116	0	0	-3.5	3.5	.0	.0
3 C HEMMSOTH	94.0	.9	.3	89	96	0	0	.3	-.3	.0	.0
4 B VOIERS	92.0	2.7	2.1	133	112	0	0	15.6	-15.6	.0	.0
5 S NEELY	93.0	.1	.0	108	107	0	0	5.0	-5.0	.0	.0
6 B LOVE	93.2	3.8	4.7<.05	90	119	0	0	-10.2	10.2	.0	.0

ERRORS FOR LISTENERS BY ATTRIBUTES

	ATTRIBUTE PRESENT								ATTRIBUTE ABSENT							
L#)	VOIC	NASL	SUST	SIBL	GRAV	COMP	EXPL	TOTAL	VOIC	NASL	SUST	SIBL	GRAV	COMP	EXPL	TOTA
1)	12	3	7	6	26	14	2	68	13	5	26	2	43	0	0	89
2)	17	8	60	13	28	1	1	127	21	5	48	4	9	2	14	89
3)	2	2	16	7	62	5	3	94	10	1	22	3	38	17	2	91
4)	1	1	95	3	36	4	0	140	24	1	32	4	40	4	7	105
5)	13	2	55	13	62	3	3	146	7	1	17	1	31	10	0	67
6)	15	10	32	3	65	4	6	129	42	3	20	4	6	5	2	80
OTAL	60	26	265	45	279	31	15	706	117	16	165	18	167	38	25	521
	ATTRIBUTE BIAS								ATTRIBUTE TOTAL							
1)	-1	-2	-19	4	-17	14	2	-21	25	8	33	6	69	14	2	157
2)	-4	3	12	9	19	-1	-13	38	38	13	108	17	37	3	15	216
3)	-8	1	-6	4	24	-12	1	3	12	3	38	10	100	22	5	185
4)	-23	0	63	-1	-4	0	-7	35	25	2	127	7	76	8	7	245
5)	6	1	38	12	31	-7	3	81	20	3	72	14	93	13	3	215
6)	-27	7	12	-1	59	-1	4	49	57	13	52	7	71	9	8	209
OTAL	-57	10	100	27	112	-7	-10	185	177	42	430	63	446	69	40	1227

ERRORS FOR EACH ITEM

ITEMS 1-28				ITEMS 29-56				ITEMS 57-84				ITEMS 85-112			
0	5	3	31	5	24	3	4	66	0	2	0	9	0	24	1
2	1	2	0	5	4	1	5	1	3	3	3	4	7	1	0
27	31	1	14	57	9	73	18	65	26	3	8	2	18	65	13
10	2	0	2	8	5	3	3	1	11	1	0	6	7	3	1
22	6	6	13	47	10	20	26	6	96	23	6	3	25	132	5
0	7	3	1	20	15	3	4	2	3	0	3	0	0	4	4
1	7	4	1	3	2	1	0	2	5	1	2	4	3	3	1